

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-200191
 (43)Date of publication of application : 04.08.1995

E5874

(51)Int.Cl.

G06F 3/06
 G06F 3/06
 G06F 3/06

(21)Application number : 06-000722
 (22)Date of filing : 10.01.1994

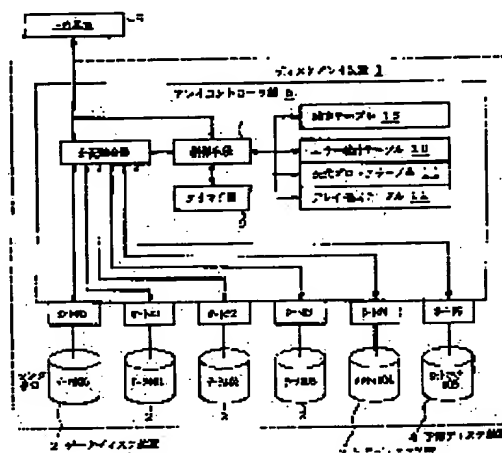
(71)Applicant : FUJITSU LTD
 (72)Inventor : SATO KEIICHI

(54) DISK ARRAY DEVICE

(57)Abstract:

PURPOSE: To quickly recognize the need for data restoration to a stand-by disk device and restore data without delaying processing from a host device

CONSTITUTION: The disk array device 1 is equipped with plural disk devices 2, and at least one redundant disk device 3 and the stand-by disk device 4. A timer means 9 measures the time of data arrival from the respective data disk devices 2 and when the data arrival is delayed exceeding a specific time, a control means 7 restores delayed data from remaining disk devices. If a data block which can not be corrected is detected on the disk device 2 or 3, data are restored from the remaining disk devices which belong to the same parity group and allocated to an alternative area. Further, data are restored to the stand-by device as to a disk device which is high in the possibility of an alternative area overflow, rotation synchronism trouble, and a fault.



LEGAL STATUS

[Date of request for examination]

10.08.2000

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2000 Japanese Patent Office

JP#09

E5874

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平7-200191

(43) 公開日 平成7年(1995)8月4日

(51) Int.Cl. ⁸	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 3/06	5 4 0			
	3 0 5 C			
	3 0 6 B			

審査請求 未請求 請求項の数13 O L (全 32 頁)

(21) 出願番号 特願平6-722
(22) 出願日 平成6年(1994)1月10日

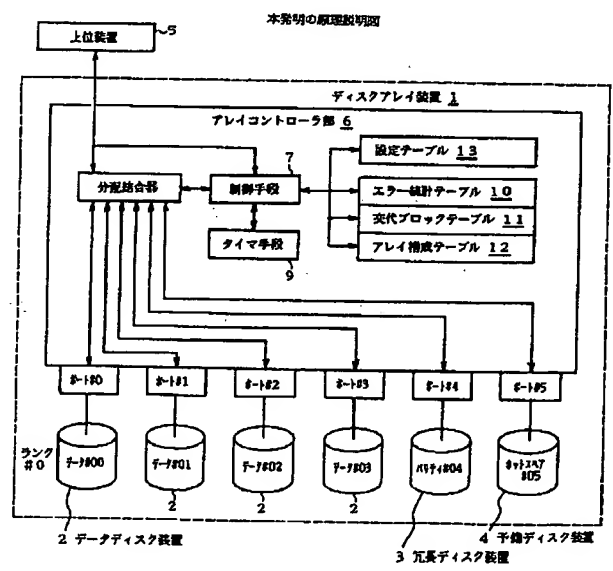
(71) 出願人 000005223
富士通株式会社
神奈川県川崎市中原区上小田中1015番地
(72) 発明者 佐藤 恵一
山形県東根市大字東根元東根字大森5400番
2 株式会社山形富士通内
(74) 代理人 弁理士 竹内 進 (外1名)

(54) 【発明の名称】 ディスクアレイ装置

(57) 【要約】

【目的】 予備ディスク装置に対するデータ復元の必要性をいち早く認識し、また上位装置からの処理を遅延することなしにデータ復元を可能とする。

【構成】 ディスクアレイ装置1は、複数のデータディスク装置2、少なくとも1台の冗長ディスク装置3及び予備ディスク装置4を備える。各データディスク装置2からのデータ到着時間をタイマ手段9で計測し、所定時間を越えて遅延した場合に、制御手段7で残りのディスク装置から遅延しているデータ復元する。ディスク装置2、3のいずれかで訂正不可能なデータブロックが検出された場合、同一パリティグループに属する残りのディスク装置からデータ復元して交代領域に割り付ける。更に、交代領域オーバフロー、回転同期障害、故障の危険性の高いディスク装置につき、予備ディスク装置へのデータ復元を行う。



【特許請求の範囲】

【請求項 1】複数のデータディスク装置 (2)、少なくとも 1 台の冗長ディスク装置 (3) 及び予備ディスク装置 (4) を備えたディスクアレイ装置に於いて、配下の各データディスク装置 (2) からのデータ到着時間を監視するタイマ手段 (9) と、

該タイマ手段 (9) で監視しているデータ到着時間が所定時間を越えて遅延した場合に、遅延したディスク装置のデータを残りのディスク装置から復元する制御手段

(7) と、を設けたことを特徴とするディスクアレイ装置。

【請求項 2】請求項 1 記載のディスクアレイ装置において、前記制御手段 (7) は、配下の各データディスク装置 (2) で訂正不可能なデータブロックが検出された場合に、訂正不可能となったデータを残りのディスク装置から復元し訂正不可能なデータブロックを検出したディスク装置の交代領域に割り付けることを特徴とするディスクアレイ装置。

【請求項 3】請求項 1 記載のディスクアレイ装置において、前記制御手段 (7) は、配下のデータディスク装置 (2) 及び冗長ディスク装置 (3) のいずれかで交代領域のブロックが全て使用された場合に、該ディスク装置のデータを予備ディスク装置 (4) に復元すると共に、予備ディスク装置 (4) に対する復元終了まで又は復元中に再度交代割付処理が要求されるまで、当該ディスク装置を論理ディスク装置の構成要素として上位装置 (5) からの入出力要求を処理することを特徴とするディスクアレイ装置。

【請求項 4】請求項 1 記載のディスクアレイ装置において、前記制御手段 (7) は、配下のデータディスク装置 (2) 及び冗長ディスク装置 (3) のいずれかで使用可能な交代領域内ブロック数が閾値まで減少した場合に、該ディスク装置のデータを予備ディスク装置 (4) に復元すると共に、データ復元終了まで又は復元中に再度交代割付処理が要求されるまで又は交代ブロックが全て使用されるまで、当該ディスク装置を論理ディスク装置の構成要素として前記上位装置 (5) からの入出力要求を処理することを特徴とするディスクアレイ装置。

【請求項 5】請求項 1 記載のディスクアレイ装置において、前記制御手段 (7) は、回転同期制御を行っている配下のデータディスク装置 (2) 及び冗長ディスク装置 (3) のいずれかの同期回転異常を検出した場合に、該ディスク装置のデータを予備ディスク装置 (4) に復元すると共に、データ復元終了まで又は復元中にライト命令を受けるまで、当該ディスク装置を論理ディスク装置の構成要素として前記上位装置 (5) からの入出力要求を処理することを特徴とするディスクアレイ装置。

【請求項 6】請求項 1 記載のディスクアレイ装置において、前記制御手段 (7) は、回転同期制御を行っている配下のデータディスク装置 (2) 及び冗長ディスク装置

(3) のいずれかの同期回転異常を検出した場合に、該ディスク装置のデータを予備ディスク装置 (4) に復元すると共に、データ復元終了まで当該ディスク装置を論理ディスク装置の構成要素として処理し、更に、データ復元中に上位装置 (5) よりライト命令を受けた場合にも、回転同期を維持することのできなくなったディスク装置を論理ディスク装置の構成要素として前記上位装置 (5) からのライト動作を行わせることを特徴とするディスクアレイ装置。

【請求項 7】請求項 1 記載のディスクアレイ装置において、前記制御手段 (7) は、配下の各ディスク装置のエラー回数とその頻度を管理するエラー統計テーブル (10) を有し、該エラー統計テーブル (10) を参照して故障の危険性の高いディスク装置を認識することを特徴とするディスクアレイ装置。

【請求項 8】請求項 7 記載のディスクアレイ装置において、前記制御手段 (7) は、前記エラー統計テーブル (10) を参照して故障の危険性の高いディスク装置を認識した場合に、故障する以前に該ディスク装置のデータを先行して前記予備ディスク装置 (4) に復元すると共に、データ復元終了まで当該ディスク装置を論理ディスク装置の構成要素として前記上位装置 (5) からの入出力要求を処理することを特徴とするディスクアレイ装置。

【請求項 9】請求項 1 記載のディスクアレイ装置において、前記制御手段 (7) は、該ディスク装置のデータを予備ディスク装置 (4) に復元すると共に、データ復元終了まで又はデータ復元中に前記上位装置 (5) から交代ブロック領域のデータブロックに対するライト命令を受けるまで、当該ディスク装置を論理ディスク装置の構成要素として前記上位装置 (5) からの入出力要求を処理することを特徴とするディスクアレイ装置。

【請求項 10】請求項 1 記載のディスクアレイ装置において、前記制御手段 (7) は、回転同期制御を行っている配下の各ディスク装置の同一トラック又は同一シリンダ位置に設けた交代領域の使用済みブロック数を管理する交代ブロックテーブル (11) を有し、前記交代ブロックテーブル (11) で管理される交代ブロック数が閾値を越え交代領域割付けのシーク動作に伴いデータ到達が遅延するディスク装置を認識した場合、該ディスク装置のデータを予備ディスク装置 (4) に復元すると共に、データ復元終了まで当該ディスク装置を論理ディスク装置の構成要素として前記上位装置 (5) からの入出力要求を処理することを特徴とするディスクアレイ装置。

【請求項 11】請求項 1 記載のディスクアレイ装置において、前記制御手段 (7) は、配下の各ディスク装置の固有情報、動作状態、物理的な設置位置を管理するアレイ構成テーブル (12) を有し、前記アレイ構成テーブル (12) を参照して配下のディスク装置のデータが予

備ディスク装置(4)に復元されていることを認識した場合に、前記予備ディスク装置(4)と故障ディスク装置との物理的位置を交換移動を促すメッセージを外部に出力して再構成させることを特徴とするディスクアレイ装置。

【請求項12】請求項1記載のディスクアレイ装置において、前記制御手段(7)は、配下の各ディスク装置の固有情報、動作状態、物理的な設置位置を管理するアレイ構成テーブル(12)を有し、前記アレイ構成テーブル(12)を参照して配下のディスク装置のデータが予備ディスク装置(4)に復元されていることを認識した場合に、前記予備ディスク装置(4)を故障ディスク装置との論理的位置を移動させ、物理的に移動することなく再構成させることを特徴とするディスクアレイ装置。

【請求項13】請求項1記載のディスクアレイ装置において、前記制御手段(7)は、前記上位装置(5)からの命令で、データディスク装置(2)及び冗長ディスク装置(3)のデータを前記予備ディスク装置(4)に復元させることを特徴とするディスクアレイ装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、アレイ状に配置された複数の複数のディスク装置を並列的にアクセスするディスクアレイ装置に関し、特に、ディスク装置の故障に対し予備ディスク装置を利用してデータ復元を行うディスクアレイ装置に関する。現在、電子計算機の外部記憶装置には、主に磁気ディスクが使用されている。

【0002】近年、電子計算機の性能向上はめざましく、外部記憶装置にも高性能のものが求められている。特に、画像処理や科学技術計算の分野では、磁気ディスク装置を並列に配置し、複数台の磁気ディスクから同時にデータを読み書きしてデータ転送を高速化したディスクアレイ装置が用いられている。また冗長ディスク装置を装備し、データディスク装置に障害が起きても残りのディスク装置からデータを復元することが可能であるが、障害に対し迅速に対応可能な信頼性の高いディスクアレイ装置が求められている。

【0003】

【従来の技術】従来、ディスクアレイ装置は小型のディスク装置を多量に使用し、システムの故障率を下げるために一般的に冗長性を持たせている。図21は従来のディスクアレイ装置を示す。ディスクアレイ装置1はアレイコントローラ部6と複数のデータディスク装置2-0～2-3、冗長ディスク装置3およびホットスペアとして知られた予備ディスク装置4で構成される。各ディスク装置2-0～2-3、3、4はアレイコントローラ部6に設けたポート15-0～15-5に接続され、並列的に動作できる。アレイコントローラ部6には、プロセッサ8と分配結合器8が設けられる。

【0004】上位装置5からのライトコマンドの発行に

伴って転送されてきたブロックデータは、アレイコントローラ部6の分配結合器8で、4台のデータディスク装置2-0～2-3のデータに分配(ストライピング)され、ポート15-0～15-3を介して並列的に書込まれる。同時に4つの分配データからパリティデータが生成され、ポート15-4を介して冗長ディスク装置3に書込まれる。

【0005】ここでブロックデータはビット又はバイト単位に分配され、また冗長ディスク装置の位置を固定していることから、いわゆるRAID3に従った制御を行っている。データのリード時には、ライト時と逆にデータディスク装置2-0～2-3からデータを読み出してアレイコントローラ部6の分配結合器8で結合して元のブロックデータを復元し、同時に冗長ディスク装置3から読み出したパリティデータと比較して正常に結合されている場合に、復元したブロックデータを上位装置5に転送する。

【0006】またディスクアレイ装置1では、仮に1台のデータディスク装置、例えばデータディスク装置2-3が故障した場合、残りのデータディスク装置2-0～2-2のデータと冗長ディスク装置3のパリティデータから故障したデータディスク装置2-3のデータを復元することができる。このため予備ディスク装置4が設けられ、例えばディスク装置2-3で致命的な故障が起きた場合、残りのデータディスク装置2-0～2-2のデータと冗長ディスク装置3のパリティデータから故障したデータディスク装置2-3のデータを復元して予備ディスク装置4に書込み、予備ディスク装置4の復元データを使用可能とする。

【0007】

【発明が解決しようとする課題】しかしながら、このような従来のディスクアレイ装置にあっては、配下のディスク装置に訂正不可能な媒体エラーや致命的なハードウェアエラーが起きた場合にのみ、故障と判断して予備ディスク装置に対するデータ復元を行っており、復元処理を行っている間、上位装置からアクセスができない不都合があった。

【0008】また、故障したディスク装置のデータを予備ディスク装置4に復元した後、故障ディスク装置を正常なものと交換し、交換したディスク装置に予備ディスク装置4から復元データをコピーした後に正常な運用に移行させていたため、データ復元やディスク装置の交換に手間と時間がかかるという問題があった。本発明は、このような従来の問題点に鑑みてなされたもので、予備ディスク装置に対するデータ復元の必要性をいち早く認識し、また上位装置からの処理を遅延することなしにデータ復元を可能とする冗長性が高く予備ディスク装置を効率良く使用可能なディスクアレイ装置を提供することを目的とする。

【0009】

【問題点を解決するための手段】図1は本発明の原理説明図である。まず本発明は、複数のデータディスク装置2、少なくとも1台の冗長ディスク装置3及び予備ディスク装置4を備えたディスクアレイ装置1を対象とする。

【タイマ監視によるデータ復元処理】このようなディスクアレイ装置1につき本発明は、配下の各データディスク装置2からのデータ到着時間をタイマ手段9で計測し、データ到着時間が所定時間を越えて遅延した場合に、制御手段7で遅延したディスク装置のデータを残りのディスク装置から復元する。

【0010】尚、冗長ディスク装置3からのデータが遅延しても、データディスク装置2のデータのみからホストデータは生成できるため、冗長ディスク装置3のパリティデータを復元する必要はない。

【データ復元を伴う交代割付処理】またディスクアレイ装置1の制御手段7は、配下の各データディスク装置2および冗長ディスク装置3のいずれかで訂正不可能なデータブロックが検出された場合に、訂正不可能となったデータを同一パリティグループに属する残りのディスク装置から復元し、訂正不可能なデータブロックを検出したディスク装置の交代領域に割り付ける。

【0011】【交代領域オーバフローに伴う予備データ復元処理】さらにディスクアレイ装置1の制御手段7は、交代領域が全て使用済みとなった場合にも予備ディスク装置へのデータ復元を行う。即ち、配下のデータディスク装置2及び冗長ディスク装置3のいずれかで交代領域が全て使用された場合に、そのディスク装置のデータを予備ディスク装置4に復元する。

【0012】この場合、予備ディスク装置4に対する復元終了まで、又は、データ復元中に再度交代割付処理が要求されるまでは、当該ディスク装置を論理ディスク装置の構成として上位装置5からの入出力要求を処理する。その後に全交代領域が使用済みとなったディスク装置は、故障ディスク装置として扱われ、論理ディスク装置の構成要素から除外される。

【0013】【交代領域オーバフロー予測に基づく予備データ復元処理】交代領域の使用済みに基づく予備ディスク装置4へのデータ復元処理は、予測的に行ってもよい。即ち、データディスク装置2および冗長ディスク装置3のいずれかで使用可能な交代領域内ブロック数（空き数）が閾値まで減少した場合に、当該ディスク装置のデータを予備ディスク装置4に復元する。

【0014】この場合、データ復元終了まで、又はデータ復元中に再度交代処理が要求されるまで又は交代ブロックが全て使用されるまで、当該ディスク装置を論理ディスク装置の構成要素として上位装置5からの入出力要求を処理する。その後は、制御手段7において故障ディスク装置として扱われる。

【同期回転異常時の予備データ復元処理】一方、ディス

クアレイ装置1の制御手段7は、配下のデータディスク装置2及び冗長ディスク装置3のいずれかでの同期回転異常を検出した場合に、該ディスク装置のデータを予備ディスク4に復元する。

【0015】この場合、データ復元終了まで又はデータ復元中にライト命令を受けるまで、当該ディスク装置を論理ディスク装置の構成要素として上位装置5からの入出力要求を処理する。その後は、制御手段7において故障ディスク装置として扱われる。この場合、データ復元中に上位装置5よりライト命令を受けても故障ディスク装置とせず、回転同期を維持することのできなくなったディスク装置をそのまま使用してライト動作を行わせるようにしてもよい。

【0016】【障害発生の予測処理】ディスクアレイ装置1の制御手段7は、ディスク装置の障害発生を予測して予備ディスク装置4に対するデータ復元を行う。即ち、制御手段7は、配下の各ディスク装置のエラー回数とその頻度を管理するエラー統計テーブル10を有し、エラー統計テーブル10を参照して故障の危険性の高いディスク装置を認識する。

【0017】【障害発生の予測に基づく予備データ復元処理】エラー統計テーブル10を参照して故障の危険性の高いディスク装置を認識した場合、故障する以前に当該ディスク装置のデータを先行して予備ディスク装置4に復元する。この場合、データ復元終了まで当該ディスク装置を論理ディスク装置の構成要素として上位装置5からの入出力要求を処理する。

【0018】データ復元が済むと、故障する危険性の高いディスク装置は、制御手段7により故障ディスク装置として扱われる。

【交代ブロックの増加に伴うディスク装置の予備データ復元処理】またディスクアレイ装置の制御手段7は、回転同期制御を行っている配下の各ディスク装置の同一トラック又は同一シリンダの交代領域の使用状態（交代ブロック数）を管理する交代ブロックテーブル11を有し、交代ブロック数が閾値に達して別のトラック又はシリンダに交代領域が割付られ、このような交代領域へのデータ割付けに伴うシーク動作でデータ到着時間が遅延する。このようにデータ到達の遅延が予測されるディスク装置のデータを予備ディスク装置4に復元する。

【0019】この場合、データ復元終了まで、又はデータ復元中に上位装置5から交代ブロック領域のデータブロックを含むライト命令を受けるまで、当該ディスク装置を論理ディスク装置の構成要素として上位装置5からの入出力要求を処理する。また、データ復元中に上位装置5から交代領域のデータブロックを含むライト命令を受けても当該ディスク装置を故障ディスク装置として除外せず、データ復元終了まで論理ディスク装置の構成要素として上位装置5からの入出力要求の処理を継続させてもよい。

【0020】〔人的介入を伴う再構成処理〕ディスクアレイ装置1の制御手段7は、配下の各ディスク装置の固有情報、動作状態、物理的な設置位置を管理するアレイ構成テーブル12を有し、アレイ構成テーブル12を参照して配下のディスク装置のデータが予備ディスク装置4に復元されていることを認識した場合に、予備ディスク装置4と故障ディスク装置との物理的位置を交換移動を促すメッセージを例えば上位装置5に送って外部に出力し、ユーザや保守要員に再構成させる。

【0021】〔人的介入を伴わない再構成処理〕ディスクアレイ装置1の制御手段7は、配下の各ディスク装置の固有情報、動作状態、物理的な設置位置を管理するアレイ構成テーブル12を参照してデータが予備ディスク装置4に復元されていることを認識した場合に、予備ディスク装置4を故障ディスク装置の論理的位置に移動させ、物理的に移動することなく再構成させる。

【0022】

〔作用〕このような本発明のディスクアレイ装置においては、次の作用が得られる。

〔タイマ監視によるデータ復元処理〕タイマ手段でデータ到達時間の監視を行い、故障を起こした1台のデータディスク装置のディスク転送が規定時間以上遅れた時、残りのデータディスク装置と冗長ディスク装置から遅延したデータディスク装置のデータを復元することによって、遅延しているデータディスク装置が各種リトライ処理を行っているような場合でも、その結果を待つことなしに上位装置に対しホストデータを転送することができ、上位装置に対する転送速度が向上する。

【0023】〔データ復元を伴う交代割付処理〕訂正不可能なデータブロックと同一のパーティグループを構成する残りのディスク装置のデータから、訂正不可能なブロックのデータを復元して交代領域に割付けることで、訂正不可能ブロックを検出したディスク装置を故障ディスクとして扱わなくとも良いため、ディスク装置を有効に利用することができる。

【0024】〔交代領域オーバーフローに伴う予備データ復元処理〕交代領域が使用済みとなっているディスク装置のデータを予備ディスク装置に復元し、故障ディスク装置と見做して交換可能とする。この場合、データ復元終了まで、又は、このディスク装置においてデータ復元中に再度交代割付処理が必要となるまで継続して使用し続けることによって、データ復元中に他のデータディスク装置または冗長ディスク装置が故障した場合にも、論理ディスク装置の冗長性を確保でき、ユーザデータを保証することができる。

【0025】〔交代領域オーバーフロー予測に基づく予備データ復元処理〕一方、交代領域の使用済みに伴う予備ディスク装置のデータ復元中に再度交代割付処理が必要となった場合、ディスク装置の冗長性が失われてしまうのに対し、交代領域の使用可能な残りブロック数が閾値

以下に減少した時に予備ディスク装置へのデータ復元を開始することで、データ復元中にも残り数分の交代割付処理かできるので、より高い信頼性でユーザデータを保証することができる。

【0026】また、データ復元中に起こり得る交代割付処理以上に残りブロック数の閾値を上げておくことで、データ復元後まで論理ディスク装置の冗長性を保つことができる。

〔同期回転異常時の予備データ復元処理〕回転同期が維持できなくなったディスク装置のデータを予備ディスク装置に復元し、故障ディスク装置として交換可能とする。

【0027】この場合、同期回転が維持できなくなったディスク装置を、データ復元終了まで、又は上位装置からライト命令を受け取るまで継続して使用し続けることによって、データ復元中に他のデータディスク装置または冗長ディスク装置が故障した場合にも、ディスク装置の冗長性を確保でき、ユーザデータを保証することができる。また、データ復元中の上位装置5からのライト命令を高速に処理できる。

【0028】一方、データ復元中にライト命令に対し、同期回転が維持できなくなっているディスク装置にライト動作を行わせてもよい。これにより回転同期を維持している時よりライト動作に時間はかかるが、回転同期を維持することが不可能となったディスク装置を、データ復元終了まで継続して使用し続けて冗長性が確保され、より高い信頼性でユーザデータを保証することができる。

【0029】〔障害発生の予測処理〕エラー統計テーブルの内容を設定されている閾値と比較することによって、故障する危険性の高いディスク装置を事前に把握できる。

〔障害発生の予測に基づく予備データ復元処理〕故障する危険性の高いディスク装置が致命的な故障を起こす以前に予備ディスク装置にデータを復元することができる。また、データ復元中に他のデータディスク装置または冗長ディスク装置が故障した場合にも、ディスクの冗長性を確保できるためユーザデータを保証することができる。

【0030】〔交代ブロック数に伴うディスク装置の予備データ復元処理〕同一のトラック又は同一シリンダの交代領域の交代ブロック数が増加すると他のトラック又はシリンダに交代領域が拡張され、このような交代ブロックをリードするとシーク動作に時間がかかり、上位装置へのデータ転送速度が遅延する。このようなディスク装置のデータを予備ディスク装置に復元し、交代領域を使用せずにリード可能な状態にすることで、データ転送の遅延を抑えることができる。

【0031】〔人的介入を伴う再構成処理〕何らかの原因により配下のディスク装置のデータが予備ディスクに復元されている場合に、予備ディスク装置のデータを

交換したディスク装置にコピーすること無しに、ユーザや保守要員が物理的な位置を移動するだけで論理ディスク装置を再構成して使用することができる。

〔人的介入を伴わない再構成処理〕 予備ディスク装置の物理的な位置を移動せずに、論理ディスク装置を再構成して使用でき、ディスク装置の交換に伴う人為的ミス防止できる。

【0032】

【実施例】

<目次>

1. ハードウェア構成と機能
2. タイマ監視によるデータ復元
3. データ復元を伴う交代割付処理
4. 交代領域オーバーフローに伴う予備へのデータ復元
5. 交代領域オーバーフロー予測に基づく予備へのデータ復元
6. 同期回転異常時の予備へのデータ復元
7. 障害発生の予測と予備へのデータ復元
8. 同一トラック内の交代ブロック数の増加に伴う予備へのデータ復元
9. 人的介入を伴う再構成
10. 人的介入を伴わない再構成

1. ハードウェア構成と機能

図2は本発明のディスクアレイ装置のハードウェア構成と機能を示した実施例構成図である。

【0033】 図2において、本発明のディスクアレイ装置1は例えば4台のデータディスク装置2-0~2-3、1台の冗長ディスク装置3、および1台の予備ディスク装置4を備え、各ディスク装置をアレイコントローラ部6に設けられたポート15-0~15-5のそれぞれに接続している。ディスクアレイ装置1における各ディスク装置は、その論理的な位置および物理的な位置がポート番号#0~#5とランク番号#0~#nで特定される。この実施例にあつては、1ランク構成を例にとっているが、必要に応じて複数ランク設けることができる。

【0034】 この実施例にあつては、ポート番号#0~#3のポート15-0~15-3にデータディスク装置2-0~2-3を接続し、またポート番号#4のポート15-4にパリティデータを格納する冗長ディスク装置3を接続している。データディスク装置2-0~2-3および冗長ディスク装置3のアドレス（デバイスID）はポート番号とランク番号で特定され、#00、#01、#02、#03、#04となる。

【0035】 4台のデータディスク装置2-0~2-3と1台の冗長ディスク装置3により、1つのパリティグループを構成している。例えば、上位装置5からのデータブロックを書き込む場合には、転送されたデータブロックをアレイコントローラ部6で4つのデータディスク装置2-0~2-3に分配するストライピング処理を行い、同時に各ストライピングデータからパリティデータ

を生成し、各ストライピングデータおよびパリティデータを並列的にポート15-0~15-4からデータディスク装置2-1~2-3および冗長ディスク装置3に供給して、並列的に書き込む。

【0036】 一方、上位装置5からのリード要求に対しては、アレイコントローラ部6でリードブロックのアドレスからデータディスク装置2-0~2-3の各アドレスを生成し、更に冗長ディスク装置3についてもアドレスを生成し、ストライピングデータおよびパリティデータを系列的にリードする。そしてアレイコントローラ部6において、リードした各ストライピングデータからリードブロックを生成し、生成したリードブロックからパリティを生成して冗長ディスク3からリードしたパリティデータとの整合をチェックし、正常であれば上位装置5にリードブロックを転送することになる。

【0037】 このようなディスクアレイ装置1におけるライト動作およびリード動作のため、アレイコントローラ部6には制御手段としてのプロセッサ7および分配結合器8が設けられている。なお、この実施例にあつては、冗長ディスク装置3を固定的に決めていることから、データブロックをビット単位あるいはバイト単位に分配して系列的に読み書きする、いわゆるRAID3として知られたディスクアレイの制御形式を例にとっている。

【0038】 更に本発明のアレイコントローラ部6にあつては、タイマ9、エラー統計テーブル10、交代ブロックテーブル11、アレイ構成テーブル12および閾値設定テーブル13を新たに設けている。プロセッサ7はタイマ9、エラー統計テーブル10、交代ブロックテーブル11、アレイ構成テーブル12および閾値設定テーブル13を使用し、配下のデータディスク装置2-0~2-3および冗長ディスク装置3のいずれかで障害が発生したことを認識すると、障害発生ディスク装置を除く他のディスク装置からのデータに基づく復元処理、あるいは障害と判定されたディスク装置のデータを予備ディスク装置4に復元する復元処理を実行する。

【0039】 即ち、本発明によるプロセッサ7の制御処理は、次に列挙する内容となる。

- ①タイマ9の監視によるリードブロックデータの復元
- ②データ復元を伴う交代割付処理
- ③交代領域がオーバーフローになった場合の予備ディスク装置4へのデータ復元
- ④交代領域のオーバーフローを予測して、予備ディスク装置4へのデータ復元
- ⑤回転同期制御を行っている各ディスク装置のいずれかにおける同期回転異常の際の予備ディスク装置4へのデータ復元
- ⑥各ディスク装置の障害の発生を予測して、予備ディスク装置4へのデータ復元
- ⑦交代ブロック数が増加したディスク装置の予備ディス

ク装置へのデータ復元

⑧予備ディスク装置4へのデータ復元が済んだ後の人的介入による再構成

⑨予備ディスク装置への復元が済んだ後の、人的介入を伴わない再構成

以下、前記①～⑨のそれぞれについて詳細に説明する。

2. タイマ監視によるデータ復元

図3は上位装置5からのリード命令に対し、アレイコントローラ部6のプロセッサ7が適当なタイミングでタイマ9を初期化して起動し、パリティグループを構成するデータディスク装置2-0～2-3のいずれか1台からのデータ転送が閾値設定テーブル13に設定されているタイムアウト時間以上遅れた場合、既に得られている残りのデータディスク装置からのデータに基づき、遅延したデータディスク装置のデータを復元し、リードブロックを生成して上位装置5に転送するようにしたことを特徴とする。

【0040】図3のフローチャートについて詳細に説明すると次のようになる。まずステップS1で、ホストコンピュータ5よりプロセッサ7がリード要求を受領すると、ステップS2で、ホストコンピュータにおける論理ディスクを構成する配下のデータディスク装置2-0～2-3および冗長ディスク装置3に対しリードコマンドを発行する。

【0041】このリードコマンドの発行に基づきステップS3で、配下のデータディスク装置2-0～2-3および冗長ディスク装置3のリード処理が行われ、リードデータが転送されてくる。このときプロセッサ7はどのディスク装置が分配結合器8に対しデータ転送を完了したか否かを監視しており、パリティデータとデータ復元に必要なリードデータが分配結合器8に転送されるか否か、ステップS4でチェックしている。

【0042】即ち、この実施例にあっては、4台のデータディスク装置2-0～2-3を配下にもつことから、パリティデータと4台のディスク装置のうちの3台のリードデータが得られると、残り1つのリードデータを復元することができるため、4台のデータディスク装置2-0～2-3のうちの3台のデータ転送と冗長ディスク装置3からのパリティデータの転送が済んだ時点で、ステップS4のデータ復元に必要なデータ受領と判定する。

【0043】続いてステップS5のデータ遅延のフラグセットをチェックするが、初期状態でフラグはリセットされていることからステップS6に進み、プロセッサ7はタイマ9の初期化および起動を行う。ステップS6におけるタイマ9の起動後、プロセッサ7は閾値設定テーブル13より、予め定められたタイムアウト時間（閾値時間）を読み出す。

【0044】ステップS8で、設定されたタイムアウト時間内に遅延しているディスク装置からのデータを受領

した場合には、分配結合器8で全てのデータディスク装置2-0～2-3からのリードデータからホストコンピュータ5に対するリードブロックデータを生成し、冗長ディスク装置3からのパリティデータとの整合性を確認した上で、ホストコンピュータ5に対しリードデータから生成されたホストブロックデータを転送する。

【0045】一方、ステップS8の設定タイムアウト時間内に残りのデータディスク装置からのリードデータを受け取ることができなかった場合には、ステップS9に進んで、データ遅延が発生したことを示すフラグをセットし、既に分配結合器8に転送されている3台のデータディスク装置からのリードデータと冗長ディスク装置3からのパリティデータを使用して、ステップS10でホストデータブロックを生成してホストコンピュータ5に転送する。

【0046】続いてプロセッサ7は、データ転送が遅延しているデータディスク装置をその後も監視し、遅延の原因と結果を示すステータス情報の受領を待つ。このステップS11におけるステータス情報の受領待ちの際に、ステップS12で再度リード要求があるか否かチェックしている。もしステータス情報を受領する前に再度リード要求があると、ステップS2に戻り、ホストコンピュータからのリード要求に基づきステップS2～S3で配下のデータディスク装置へのリードコマンドの発行に伴うリード処理およびデータ転送を行わせ、この場合にも前回と同様、同じデータディスク装置からのデータ遅延を起こすことになる。

【0047】この場合、前回の処理で既にデータ遅延フラグがセットされていることから、ステップS5から直ちにステップS10に進み、ステップS6～S9のタイマ起動に基づく監視処理は行わず、ステップS4で分配結合器8に得られた遅延ディスク装置を除くデータディスク装置からのリードデータと冗長ディスク装置3からのパリティデータによってデータを復元してホストブロックデータを生成して、ホストコンピュータ5に直ちに転送する。

【0048】ステップS11で、遅延しているディスク装置の原因と結果を示すステータス情報が受領されると、ステップS13に進み、データ遅延を示すフラグをクリアし、ステップS14で、データ遅延の原因が各種のリトライ処理にあったか否かチェックする。遅延理由が各種のリトライ処理にあった場合には、続いてステップS15で、データ遅延の原因が訂正不可能なリードデータのリトライ処理であったか否かチェックする。

【0049】もし訂正不可能なリードデータのリトライ処理であった場合には、図4のステップS16以降の処理に進む。ここで、訂正不可能なリードデータのリトライ処理は主にディスク媒体の媒体欠陥を原因とするものであり、交代領域に対する割付処理でリカバーすることができる。一方、ステップS15でデータ遅延の原因が

訂正不可能なリードデータのリトライ処理でなかった場合、即ちデータディスク装置のエラーであった場合には、図5のステップS21に示すエラー統計テーブル10の更新処理に移行する。

3. データ復元を伴う交代割付処理

図4のフローチャートは、図3のステップS15に示したデータ遅延を起こしたデータディスク装置について訂正不可能ブロックが検出された場合の交代割付処理を示している。

【0050】図4において、ステップS1～S15の処理は図3と同じであり、ステップS2～S14については省略して示している。ステップS15でデータ遅延の原因が訂正不可能なデータブロックの検出にあったことが判別された場合、ステップS16に進み、訂正不可能なデータブロックのアドレスから他の正常なデータディスク装置での同一パリティグループを構成するアドレスを算出し、次のステップS17で、正常なデータディスク装置2および冗長ディスク装置3から訂正不可能なデータブロックが検出されたディスク装置のデータ復元に必要なデータリードのためのリードコマンドを発行する。

【0051】続いてステップS18で、正常なデータディスク装置2および冗長ディスク装置3から得られたリードデータおよびパリティデータに基づき、訂正不可能なブロックから検出されたディスク装置のリードデータを復元し、ステップS19で、データ遅延を起こしたデータディスク装置の交代領域に復元したデータを書き込む割付けを行う。

【0052】このようなステップS16～S19における訂正不可能なデータブロックを他のディスク装置からのリードデータで復元して交代領域に割り付ける処理は、ディスクアレイ装置1がホストコンピュータ5からの命令を何も実行していない空き時間に行うことが望ましい。またステップS19で、交代領域に対するデータ割付けが済んだ場合には、図5のステップS21およびS22に示すエラー統計テーブル10の更新および交代ブロックテーブル11の更新を行うようになる。

4. 交代領域オーバフローに伴う予備へのデータ復元
図5のフローチャートは、配下のディスク装置において交代割付用の交代領域が全て使い尽くされたときに、交代領域を全て使い尽くしたディスク装置のデータを予備ディスク装置4に復元するようにしたことを特徴とする。

【0053】具体的に図5について説明すると次のようになる。なお図5のフローチャートにあっては、データディスク装置2-3（機番アドレス#03）で交代領域が使い尽くされてオーバフローとなった場合の処理を示している。尚、図のフローチャート中でデータディスク装置は、データ#03と省略して示している。まずステップS20で、配下のディスク装置のうちのデータディ

スク装置2-3が訂正不可能なブロックの検出に基づき、交代領域への割付処理を実施したとする。続いてステップS21で、交代処理の原因となったエラーのエラー統計テーブル10に対する更新を行い、またステップS22で、交代ブロックテーブル11に対する更新を行う。

【0054】次にステップS23でデータディスク装置2-3に関する交代ブロックテーブル11の内容を参照し、交代領域の残りブロック数が0となって使い尽くされていた場合には、ステップS24に進み、プロセッサ7は直ちに交代領域が使い尽くされたデータディスク装置2-3のデータを予備ディスク装置4に復元するための処理を開始する。

【0055】このデータ復元処理は、交代領域が使い尽くされたデータディスク装置2-3を除く他のデータディスク装置2-0～2-2および冗長ディスク装置3のリード処理を行って、分配結合器8でデータディスク装置2-3のデータを復元し、これを予備ディスク装置4に書き込む処理を行う。ステップS24で予備ディスク装置4へのデータ復元を開始すると、データ復元中におけるホストコンピュータ5からのリード要求に対してはステップS25～S30に示す処理が行われ、一方、ライト要求についてはステップS32～S38（S41、S42を含む）の処理が行われる。

【0056】まずホストコンピュータ5からデータ復元中にリード要求を受けると、このリード要求はステップS25で判別され、ステップS26に進む。ステップS26にあっては、データディスク装置2-3が故障ディスク装置に設定されているか否かチェックする。最初、データディスク装置2-3は故障ディスク装置には設定されていないことから、ステップS27に進み、データディスク装置2-3における交代割付処理が必要なリード要求か否かチェックする。

【0057】ステップS27でデータディスク装置2-3の交代割付処理を必要としないリード要求であった場合には、ステップS28に進み、全てのデータディスク装置2-0～2-3および冗長ディスク装置3に対しリードコマンドを発行してリード動作を行わせ、分配結合器8で各データディスク装置2-0～2-3のリードデータからホストブロックデータを生成し、パリティデータとの整合をチェックした後、ホストコンピュータ5に転送する。

【0058】一方、ステップS27でデータディスク装置2-3に関し、交代割付処理を必要とするリード要求であった場合には、データディスク装置2-3について交代処理を行ってリードデータを転送すると処理時間がかかることから、ステップS29でデータディスク装置2-3を故障ディスク装置に設定し、ステップS30でディスク装置2-3を除くデータディスク装置2-0～2-2および冗長ディスク装置3に対しリードコマンド

を発行して、得られたデータからホストブロックデータを復元してホストコンピュータ5に転送する。

【0059】次に図6のステップS31で予備ディスク装置4へのデータ復元中にホストコンピュータ5からのライト要求が判別されると、ステップS32に進み、データディスク装置2-3は故障ディスク装置に設定されているか否かチェックする。故障ディスク装置に設定されていなければステップS33に進み、データディスク装置2-3の交代割付処理を伴うライト要求か否かチェックする。交代割付処理を伴うライト要求でなければ、ステップS34で全てのデータディスク装置2-0~2-3および冗長ディスク装置3に対しライト命令を実行する。

【0060】一方、ステップS33でデータディスク装置2-3に関し交代割付処理を必要とするライト要求であった場合には、データディスク装置2-3の交代領域は使い尽くされて使用不可となっていることから、ステップS35に進み、データディスク装置2-3を故障ディスク装置に設定する。

【0061】続いてステップS36で、故障ディスク装置に設定されたデータディスク装置2-3を除くデータディスク装置2-0~2-2および冗長ディスク装置3に対しライト命令を実行する。この場合、故障ディスク装置に設定されたデータディスク装置2-3分の分配データ（ストライピングデータ）が欠落するが、他のディスク装置のライトデータおよびパリティデータから復元可能であることから問題はない。

【0062】なお、予備ディスク装置4に対する復元済みのデータについてのライト要求であった場合には、ライト要求の実行完了後に再度、予備ディスク装置4への復元データに割り付ける必要がある。ステップS31のライト要求に伴う処理が済むと、ステップS37で、予備ディスク装置4への復元を行っているデータディスク装置2-3以外のデータディスク装置2-0~2-2に障害発生か否かをチェックした後、ステップS38で予備ディスク装置4へのデータ復元終了を監視し、予備ディスク装置4へのデータ復元終了まで以上の処理を繰り返す。

【0063】ステップS38で予備ディスク装置へのデータ復元終了が判別されると、ステップS39で、データ復元中にデータディスク装置2-3は故障ディスク装置に設定されたか否かチェックし、もし設定されていなければステップS40でデータディスク装置2-3を故障ディスク装置とする。一方、データ復元中のステップS37で、復元対象となっているデータディスク装置2-3以外のデータディスク装置2-0~2-2または冗長ディスク装置3に障害が発生した場合には、ステップS41に進んで、データディスク装置2-3が故障ディスク装置に設定されているか否かチェックする。

【0064】もし故障ディスク装置に設定されていた場

合にはパリティグループに属する2台のディスク装置に故障が起きていることから、この場合には冗長性が失われ、ユーザデータの崩壊につながり、異常終了となる。ステップS40でデータディスク装置2-3がまだ故障ディスク装置として設定されていない場合には、ステップS42に進み、データディスク装置2-3について行っていた予備ディスク装置4へのデータ復元を中止し、新たに障害を起こしたディスク装置のデータを予備ディスク装置4に対し行うようにする。即ち、交代領域が使い尽くされた場合の予備ディスク装置4へのデータ復元に対し、致命的な障害となるような故障ディスク装置についての予備ディスク装置4へのデータ復元処理を優先させる。

【0065】図5および図6に示した交代領域オーバーフローに伴う予備ディスク装置へのデータ復元処理にあつては、交代領域が使い尽くされたデータ復元の対象となっているデータディスク装置2-3の故障ディスク装置への設定は、もしリード要求またはライト要求により交代割付処理を必要としなければ、予備ディスク装置4へのデータ復元終了まで交代領域が使い尽くされたデータディスク装置2-3を正常なディスク装置としてリード動作またはライト動作することができ、冗長性および並列アクセス性能が確保できる。

【0066】一方、データ復元中に割付交代処理を必要とするリード要求またはライト要求があると、交代領域が使い尽くされたデータディスク装置2-3は故障ディスク装置に設定され、リード動作およびライト動作の対象から除外される。しかしリード動作については、他のディスク装置のリードデータから故障ディスク装置に設定されたデータディスク装置2-3のリードデータを復元できることから、リード性能の低下は起きない。

【0067】またライト動作についても、故障ディスク装置として設定されたデータディスク装置2-3に対するライト動作が行われただけであることから、ライト性能も低下することはない。

5. 交代領域オーバーフロー予測に基づく予備へのデータ復元

図7および図8のフローチャートは、ディスク装置の交代領域の残り数が予め定めた閾値に減少したときに予備ディスク装置へのデータ復元を行うようにしたことを特徴とする。

【0068】図7において、ステップ20で例えばデータディスク装置2-3が交代割付処理を実施したとすると、ステップS21でエラー統計テーブル10の更新を行い、またステップS22で交代ブロックテーブル11の更新を行う。次にステップS23で、閾値設定テーブル13を参照して、予め定めたデータ復元処理を開始するための交代領域の残りブロック数の閾値を求め、ステップS24で、交代ブロックテーブル11から求めたデータディスク装置2-3の現在の交代領域残りブロック

数と比較する。

【0069】この交代領域残りブロック数が閾値と等しければ、ステップS25に進み、データディスク装置2-3のデータを予備ディスク装置4に復元する処理を開始する。予備ディスク装置4へのデータ復元中にホストコンピュータ5からリード要求があると、ステップS26~S32の処理が行われ、一方、ライト要求があると、図8のステップS33~S40の処理が行われる。

【0070】図7のデータ復元中のリード要求については、ステップS26でリード要求が判別されてステップS27に進み、データディスク装置2-3が故障ディスク装置に設定されていないければ、ステップS28で交代割付処理を必要とするか否かチェックし、必要としなければステップS29で、パリティグループに含まれる全ディスク装置を対象にリード命令を実行する。

【0071】ステップS28で、交代領域残りブロック数が閾値に減少したデータディスク装置2-3に関し交代割付処理が必要となった場合には、ステップS30で、交代領域残りブロック数が0に達したか否かチェックし、達していなければ、ステップS31の故障ディスク装置への設定を行わず、ステップS32で、データディスク装置2-3以外のパリティグループに属するディスク装置でリード命令を実行する。

【0072】即ち、この場合の予備ディスク装置へのデータ復元の開始は閾値分の空きブロックが残っている状態で開始していることから、交代領域残りブロック数が0になるまではデータディスク装置2-3を故障ディスク装置とはせず、冗長性を確保する。データ復元中のライト要求は図8のステップS33で判別され、ステップS34に進み、交代領域残りブロック数が閾値に減少したデータディスク装置2-3は故障ディスク装置に設定されたか否かチェックし、設定されていないければ、ステップS36で交代割付処理を必要とするライト要求か否かチェックし、必要としなければステップS37で、パリティグループを構成する全ディスク装置に対しライト命令を実行する。

【0073】ステップS36でデータディスク装置2-3に関し交代割付処理を必要とするライト要求であった場合には、ステップS38で交代領域残りブロック数が0か否かチェックし、0に達するまではステップS39でデータディスク装置2-3を故障ディスク装置に設定せず、ステップS40でデータディスク装置2-3以外のディスク装置でライト命令を実行する。

【0074】このようなライト要求に対するデータ復元中の処理が済むと、ステップS41でデータディスク装置2-3以外のディスク装置の障害発生の有無をチェックした後、ステップS40で予備ディスク装置4へのデータ復元終了の有無をチェックし、これを繰り返す。ステップS42で予備ディスク装置4へのデータ復元終了が判別されると、ステップS43で、データ復元中にデ

ータディスク装置2-3は故障ディスク装置に設定されたか否かチェックし、設定されていないければステップS44で故障ディスク装置に設定して、一連の処理を終了する。

【0075】一方、ステップS41でデータ復元中にデータディスク装置2-3以外のディスク装置で障害が発生した場合には、ステップS45に進み、このときデータディスク装置2-3が故障ディスク装置に設定されていないければ、ステップS46で予備ディスク装置4へのデータ復元を中断し、新たに故障した他のディスク装置のデータの予備ディスク装置4への復元を開始する。

【0076】またステップS45で、データ復元中に既にデータディスク装置2-3が故障ディスク装置に設定されていた場合には、同一パリティグループに含まれる2台のディスク装置に故障が起きたことから冗長性が失われ、ユーザデータの崩壊によって異常終了となる。

6. 同期回転異常時の予備へのデータ復元

図2に示した本発明のディスクアレイ装置1にあっては、アレイコントローラ部6の配下に設けたデータディスク装置2-0~2-3、冗長ディスク装置3および予備ディスク装置4は、スピンドルモータによるディスク媒体の回転制御につき回転同期をとる制御を行う場合がある。このように各ディスク装置のスピンドルモータの回転同期制御が行われている場合には、特定のディスク装置に回転同期の異常が起きると、異常を起こしたディスク装置のアクセス性能の低下に伴って全体的な性能低下を引き起こすことになる。

【0077】そこで本発明にあっては、図9のフローチャートに示すように回転同期の維持ができなくなったことを検出して、そのディスク装置のデータを予備のディスク装置に復元する処理を行うようにしたことを特徴とする。図9において、パリティグループを構成しているディスク装置の中の特定のディスク装置、例えばデータディスク装置2-3で回転同期が維持できなくなったとき、この回転同期の異常をプロセッサ7が受領し、ステップS2で、回転同期が維持できなくなったデータディスク装置2-3のデータを予備ディスク装置4に復元するデータ復元処理を開始する。

【0078】即ち、残りの回転同期が正常なデータディスク装置2-0~2-2および冗長ディスク装置3にリード命令を発行して、得られたリードデータを分配結合器8で生成して予備ディスク装置4に書き込む復元処理を開始する。予備ディスク装置4へのデータ復元中に、ステップS3でホストコンピュータ5からのリード要求が判別されると、ステップS4に進み、データディスク装置2-3以外の同一パリティグループに属するディスク装置のリード命令の実行でホストブロックデータをリードデータから生成して、ホストコンピュータ5に転送する。

【0079】またデータ復元中にステップS5でホスト

コンピュータ5からのライト要求を判別すると、ステップS6で回転同期が維持できなくなっているデータディスク装置2-3を故障ディスク装置に設定し、データディスク装置2-3を除く同一グループに属するディスク装置でライト命令を実行する。この場合、同期回転が維持できないデータディスク装置2-3はライト命令の実行対象から除外されているため、ライト命令は同期回転を維持しているディスク装置で行われ、ライト処理を高速に行うことができる。

【0080】リード要求またはライト要求に伴う処理が済むと、ステップS8で同期回転が維持できないデータディスク装置2-3以外のディスク装置に障害発生があったか否かチェックした後、ステップS9で予備ディスク装置4へのデータ復元終了の有無をチェックし、データ復元終了まで以上の処理を繰り返す。ステップS9で予備ディスク装置4へのデータ復元終了が判別されると、ステップS10で、データ復元中に同期回転が維持できなくなっているデータディスク装置2-3が障害ディスク装置に設定されたか否かチェックし、もし設定されていなければ、ステップS11で故障ディスク装置に設定する。

【0081】一方、データ復元中にステップS8で同期回転が維持できなくなったデータディスク装置2-3以外のディスク装置に障害が発生した場合には、ステップS12に進み、障害発生以前にライト命令を実行したか否かチェックし、もし障害発生以前にライト命令を実行していると、同期回転が維持できなくなったディスク装置2-3を含めて同一パリティグループに属する2台のディスク装置が故障したことから冗長性が失われ、ユーザデータの崩壊として異常終了に至る。

【0082】一方、障害発生以前にライト命令が実行されていなければステップS13に進み、現在行っている同期回転が維持できないデータディスク装置2-3に関する予備ディスク装置4へのデータ復元を中止し、新たに障害発生となった他のディスク装置のデータを予備ディスク装置4に復元する。この場合の予備ディスク装置へのデータ復元は、同期回転が維持できなくなっているデータディスク装置2-3を正常なディスク装置として扱うことから、データ復元処理に多少時間がかかることになる。更に、新たに故障したデータディスク装置の予備ディスク装置へのデータ復元が終了したならば、データ復元ができた予備ディスク装置を後から故障した障害ディスク装置と物理的に入れ替え、予備ディスク装置の位置に別の新たなディスク装置をセットして再度、同期回転が維持できていないデータディスク装置2-3のデータの復元処理を行う。

【0083】図10は同期回転を維持できなくなったディスク装置に対するデータ復元処理の第2実施例を示したフローチャートである。この第2実施例にあつては、データ復元中にホストコンピュータからのライト要求を

受けても同期回転を維持できなくなったディスク装置を故障ディスク装置に設定せず、同期回転を維持できなくなっているディスク装置を正常なディスク装置としてライト動作を実行するようにしたことを特徴とする。

【0084】即ち、図10のステップS1~S4は図9のフローチャートと同じであるが、データ復元中のライト要求をステップS5で判別すると、ステップS6で、同期回転を維持できなくなっているデータディスク装置2-3を故障ディスク装置に設定せず、全ディスク装置でライト命令を実行する。このため、ライト要求に対するライト命令の実行時間は同期回転を維持できなくなっているデータディスク装置2-3の性能低下に依存するため、ライト処理が低速で時間がかかるようになる。しかしながら、ディスクアレイ装置1としての冗長性は維持できる。

【0085】更に、データ復元中にステップS7で同期回転が維持できなくなっているデータディスク装置2-3以外のディスク装置に障害が発生したことを判別すると、ステップS10に進み、現在の予備ディスク装置4へのデータ復元を中止し、新たに障害発生となったディスク装置のデータの予備ディスク装置への復元を先行して行う。

【0086】これに対し、図9に示した実施例では、データ復元中にライト要求があった場合には、同期回転が維持できなくなったデータディスク装置2-3を故障ディスク装置に設定してしまっているため、他のディスク装置の障害発生に対し同一パリティグループに属する2台のデータディスク装置が故障となって冗長性が失われ、ユーザデータの崩壊につながっている。

【0087】しかし、図10の第2実施例では同期回転が維持できなくなってもデータディスク装置2-3を故障ディスク装置としていないことから、他のディスク装置の障害に対し冗長性を失うことなく予備ディスク装置へのデータ復元ができ、その後に同期回転が維持できなくなっているデータディスク装置2-3の予備ディスク装置4へのデータ復元が可能となる。

【0088】更に、ステップS8で予備ディスク装置へのデータ復元の終了が判別されると、ステップS9で、最終的に同期回転が維持できなくなっているデータディスク装置2-3を故障ディスク装置と設定する。

7. 障害発生の予測と予備へのデータ復元図2に示した本発明のディスクアレイ装置1にあつては、次のようなエラーを検出したときにプロセッサ7にエラー報告を行う。

- 【0089】①ディスク装置内部のリトライ処理によって回復されたリードエラー
- ②ディスク装置内部のリトライ処理によって回復されたポジショニング系のエラー
- ③ディスク装置内部のリトライ処理によって回復されたパリティエラー

④交代割付処理で回復されたライトエラー

⑤回復不可能なリードエラー

プロセッサ7はこのようなエラー検出に基づく報告を受領すると、エラー統計テーブル10を作成または内容更新すると共に、閾値設定テーブル13に予め設定されている閾値とエラー回数を比較し、エラー回数が閾値を越えているディスク装置を故障する危険の高いディスク装置として扱う障害予測処理を行う。

【0090】具体的には、図11のフローチャートのステップS1～S6により障害発生の予測処理が行われる。まずステップS1で、配下のディスク装置の各種のエラーを検出して報告を受けると、ステップS2で、エラー統計テーブル10の更新の必要性の有無をチェックし、必要があれば、ステップS3でエラー統計テーブル10の作成または更新を行う。

【0091】続いてステップS4で、閾値設定テーブル13を参照して予め定めた閾値を求め、ステップS5でエラー回数と閾値を比較する。エラー回数が閾値以上であればステップS6に進み、そのディスク装置を故障の危険性が高いディスク装置と認識する。ステップS6以降にあっては、ディスク装置2-3が故障の危険性が高いディスク装置と認識された場合を例にとって具体的に説明する。

【0092】ステップS6で、データディスク装置2-3が故障の危険性が高いディスク装置と認定されると、ステップS7に進んで、プロセッサ7は直ちに認定されたデータディスク装置2-3のデータを予備ディスク装置4に復元するデータ復元処理を開始する。データ復元中にホストコンピュータ5からのリード要求をステップS8で判別すると、ステップS9～S14に示す処理が行われる。即ちステップS9で、故障の危険性が高いと認識されたデータディスク装置2-3は故障ディスク装置に設定されているか否かチェックし、設定されていないければ、ステップS10でデータディスク装置2-3に各種のエラー発生があったか否かチェックし、エラー発生がなければステップS11で、全ディスク装置に対しリード命令を実行させて、ホストブロックデータを転送する。

【0093】ステップS10でデータディスク装置2-3に各種のエラーが発生した場合には、ステップS12で各種リカバリ処理の成功の有無をチェックし、リカバリ処理に成功すれば、ステップS13での故障ディスク装置への設定を行わず、ステップS14で、故障の危険性が高いと認識されているデータディスク装置2-3を除く他のディスク装置でリード命令を実行して、復元したホストブロックデータを上位装置5に転送する。ステップS12で各種リカバリ処理に失敗した場合には、ステップS13で、故障の危険性が高いデータディスク装置2-3は故障ディスク装置に設定され、処理対象から除外される。

【0094】データ復元中にホストコンピュータ5からのライト要求が図12のステップS15で判別されると、ステップS16～S21の処理が行われる。即ち、ステップS16で、故障の危険性が高いデータディスク装置2-3は故障ディスク装置に設定されているか否かチェックされ、設定されていないければステップS17で、各種エラーが発生したか否かチェックし、発生していなければステップS18で、全ディスク装置に対しライト命令を実行させる。

【0095】ステップS17で、故障の危険性が高いと認識されているデータディスク装置2-3で各種エラーが発生した場合には、ステップS19で、各種リカバリ処理の成功をチェックし、成功すれば、ステップS20における故障ディスク装置への設定を行わず、ステップS21で、データディスク装置2-3を除く他のディスク装置でライト命令を実行する。リカバリ処理に失敗すれば、ステップS20でデータディスク装置2-3を故障ディスク装置に設定する。

【0096】データ復元中にステップS22で、故障の危険性が高いデータディスク装置2-3以外のディスク装置に障害が発生すると、ステップS26で、データディスク装置2-3が故障ディスク装置に設定されていないければ、ステップS27で新たに障害発生となった障害ディスク装置の予備ディスク装置に対するデータ復元を先行し、その後に故障の危険性が高いデータディスク装置2-3の予備ディスク装置へのデータ復元を行う。

【0097】ステップS26で、データ復元中にデータディスク装置2-3が既に故障ディスク装置に設定されていた場合には、同一パリティグループに属する2つのディスク装置に障害が発生して冗長性が失われていることから、ユーザデータの崩壊として異常終了する。ステップS23で、予備ディスク装置4に対するデータ復元の終了が判別されると、ステップS24に進み、データディスク装置2-3が故障ディスク装置に設定されていないければ、ステップS25で故障ディスク装置への設定を行った後、一連の処理を終了する。

8. 同一トラック内の交代ブロック数の増加に伴う予備へのデータ復元

図2に示した本発明のディスクアレイ装置1にあっては、ホストコンピュータ5からのリード要求またはライト要求に対し各ディスク装置のデータブロックが媒体欠陥などにより訂正不可能なデータブロックとなった場合には、交代領域への割付処理が行われる。この割付処理の対象となる交代領域は同一トラック上または同一シリンダ上に存在するが、同一トラック上または同一シリンダ上の交代領域ブロック数が、交代処理が進んで残り数が少なくなると、別のトラックまたは別のシリンダ位置を交代領域として新たに確保するため、その後の交代領域に対するリード動作またはライト動作の際のシーク時間が長くなる。

【0098】そこで図13のフローチャートに示すように、同一トラックまたは同一シリンダに含まれる交代ブロック数が予め定めた閾値を越えたディスク装置については、シーク時間の増加に伴う性能低下の原因になることから、予備ディスク装置へのデータ復元を行う。これを図13のフローチャートについて説明すると次のようになる。

【0099】図13において、ステップS1で配下のディスク装置例えばデータディスク装置2-3で交代ブロックの割付処理が行われ、これをプロセッサ7が受領したとする。プロセッサ7は交代処理の報告を受けて、ステップS2で交代ブロックテーブル11の作成または更新を行い、続いてステップS3で、閾値設定テーブル13を参照して同一トラックまたは同一シリンダに割付け可能な交代ブロック数の閾値を読み出す。

【0100】次にステップS4で、現在の同一トラックまたは同一シリンダの交代ブロック数と閾値設定テーブル13から読み出した閾値とを比較し、閾値を越えている場合には、このデータディスク装置2-3について、ステップS5において予備ディスク装置4へのデータの復元を開始する。データ復元中にホストコンピュータ5からのリード要求がステップS6で判別すると、ステップS7~S10の処理を行う。即ちステップS7で、同一トラックまたは同一シリンダの交代ブロック数が閾値を越えたデータディスク装置2-3が既に故障ディスク装置に設定されているか否かチェックし、設定されていなければステップS8で、リード要求ブロックは交代ブロックを含んでいるか否かチェックし、含んでいなければステップS9で、全ディスク装置でリード命令を実行する。

【0101】ステップS8でリード要求ブロックに交代ブロックを含んでいた場合には交代ブロックのシーク動作に時間がかかることから、ステップS10に進み、データディスク装置2-3以外のディスク装置でリード命令を実行し、このリード命令で得られたリードデータおよびパリティデータからホストブロックデータを復元してホストコンピュータ5にデータ転送する。

【0102】予備ディスク装置4に対するデータ復元中に、図14のステップS11でホストコンピュータ5からのライト要求が判別されると、ステップS12~S16の処理が行われる。即ち、ステップS12で同一トラックまたは同一シリンダの交代ブロック数が閾値に達したデータディスク装置2-3は故障ディスク装置に設定されているか否かチェックし、設定されていなければステップS3で、ライト要求ブロックは交代ブロックを含んでいるか否かチェックする。

【0103】交代ブロックを含んでいなければステップS14で、全ディスク装置でライト命令を実行する。一方、ステップS13でライト要求ブロックにデータディスク装置2-3の交代ブロックを含んでいる場合、ステ

ップS14で同一トラックまたは同一シリンダの交代ブロック数が閾値に達したデータディスク装置2-3を故障ディスク装置に設定する。そしてステップS16で、故障ディスク装置に設定したデータディスク装置2-3を除く他のディスク装置でライト命令を実行する。

【0104】更に、データ復元中にステップS17でデータディスク装置2-3以外のディスク装置に障害が発生すると、ステップS21で、現在予備ディスク装置4へのデータ復元の対象となっているデータディスク装置2-3が故障ディスク装置に設定されていないことを条件に、ステップS22で、新たに障害発生となった障害ディスク装置の予備ディスク装置へのデータ復元を行う。

【0105】そして新たな障害ディスク装置へのデータ復元が終了すると、一旦中断したデータディスク装置2-3に関する予備ディスク装置へのデータ復元を行う。一方、データ復元中に受けたライト要求ブロックにデータディスク装置2-3の交代ブロックが含まれていた場合には、ステップS15で故障ディスク装置に設定されていることから、この場合には同一パリティグループに含まれる2台のディスク装置で故障が起きたこととなり、冗長性が失われ、ユーザデータの崩壊として異常終了となる。

【0106】ステップS18にあつては、予備ディスク装置4へのデータ復元の終了をチェックしており、データ復元を終了するとステップS19で、データディスク装置2-3の故障ディスク装置への設定が済んでいることをチェックすると、ステップS20で故障ディスク装置に設定した後、一連の処理を終了する。図15は同一トラックまたは同一シリンダの交代ブロック数が閾値に達したときの予備ディスク装置へのデータ復元処理の第2実施例を示したフローチャートである。この図15のフローチャートに示す第2実施例にあつては、予備ディスク装置へのデータ復元中にホストコンピュータから交代ブロックを含むライト要求を受けても、データ復元対象となっているデータディスク装置2-3を故障ディスク装置に設定せずに、論理ディスク装置を構成する有効なディスク装置として冗長性を確保するようにしたことを特徴とする。

【0107】即ち、図15のステップS1~S5の処理は図13と同じであるが、データ復元中にステップS6でホストコンピュータ5からのリード要求を判別した場合には、ステップS7で、全リード命令を実行する。またデータ復元中にステップS8でホストコンピュータ5からのライト要求を判別した場合には、ステップS9で、全ディスク装置でライト命令を実行する。

【0108】このため、予備ディスク装置へのデータ復元の対象となったデータディスク装置2-3は、復元終了までホストコンピュータ5の論理ディスク装置を構成する有効なディスク装置として扱われることで冗長性が

確保される。更にデータ復元中において、図16のステップS11で、データディスク装置2-3以外のディスク装置で障害が発生した場合には、ステップS15で、新たに障害発生となった障害ディスク装置のデータの予備ディスク装置4への復元を行った後、一旦中断したデータディスク装置2-3に関する予備ディスク装置4へのデータ復元を行う。

【0109】データ復元中にステップS12で予備ディスク装置4へのデータ復元の終了が判別されると、ステップS13に進む。ここで初めてデータディスク装置2-3は故障ディスク装置に設定され、論理ディスク装置を構成するディスク装置の中から除外される。

9. 人的介入を伴う再構成

図2に示した本発明のディスクアレイ装置1にあっては、何らかに原因により特定のディスク装置で予備ディスク装置へのデータ復元が必要となり、予備ディスク装置4へのデータ復元が終了すると、データ復元の済んだ予備のディスク装置をホストコンピュータの論理ディスク装置を構成するディスク装置に入れ替える再構成が必要となる。

【0110】即ち、ホストコンピュータ5から見た論理ディスク装置を構成するディスク装置の中の故障ディスク装置を取り外して、データを復元した予備ディスク装置4に置き替える処理が必要となる。図17のフローチャートにあっては、予備ディスク装置4に対するデータ復元の終了後の故障ディスク装置との入れ替えを、ユーザや保守要員などの人的介入により行うようにしたことを特徴とする。

【0111】図17のフローチャートにおいて、まずステップS1で、何らかの原因によりデータディスク装置2-3の予備ディスク装置4に対するデータの復元が必要なことを受領すると、ステップS2に進み、プロセッサ7はアレイ構成テーブル12を参照し、その種別情報から予備ディスク装置4を選択して、ステップS3で予備ディスク装置4へのデータ復元を開始する。

【0112】ステップS2のデータ復元開始前のアレイ構成テーブル12の内容は、例えば図18(B)に示すようになる。図18(B)において、アレイ構成テーブル12はディスクアレイ装置1における論理ディスクを構成する複数のディスク装置をランク番号、ポート番号、動作状態を示す種別情報で示している。即ち、アレイ構成テーブル12内において各ディスク装置は、図18(A)に示すように、「a b y」で表わされる。この内、先頭のaはランク番号、次のbはポート番号、最後のyは装置の動作状態即ち役割を示す種別情報である。この種別情報はデータディスク装置はD、冗長ディスク装置はP、予備ディスク装置はH、故障ディスク内はF、データ復元中ディスク装置はRが使用される。

【0113】図18(B)のデータ復元前のアレイ構成テーブル12にあっては、ポート番号#0~#3のディ

スク装置の論理基板は00D~03Dであり、データディスク装置2であることが判る。またポート番号#4のディスク装置は論理基板04Pとなって、冗長ディスク装置3となっていることが判る。更にポート番号#5のディスク装置は論理基板05Hから、予備ディスク装置となっていることが判る。

【0114】このような図18(B)に示すデータ復元前のアレイ構成テーブル12の参照により、ステップS3で予備ディスク装置4を認識して予備ディスク装置4へのデータ復元を開始する。予備ディスク装置4へのデータ復元中にあるのは、ステップS4でアレイ構成テーブル12の内容を更新する。即ち、図18(C)に示すように、データ復元を必要とするポート番号#3のデータディスク装置2-3の内容を「03F」と故障ディスク装置に変更し、データ復元を行っているポート番号#5の予備ディスク装置4の内容を「05R」として、データ復元中のディスク装置であることを示すように更新する。

【0115】ステップS5で予備ディスク装置5へのデータ復元が終了すると、ステップS6で、データ復元終了に伴うアレイ構成テーブル12の更新を行う。即ち、図18(D)に示すように、故障ディスク装置となったポート番号#3のディスク装置のデータの存在を消去し、ポート番号#5の予備ディスク装置4については復元したデータ#03の存在を登録すると共に、「05D」に更新してデータディスク装置となったことを示す。

【0116】続いてステップS7で、予備ディスク装置4へのデータ復元終了をホストコンピュータに通知し、ホストコンピュータ5側のディスプレイ装置などを使用して、ユーザまたは保守要員に対し予備ディスク装置を故障ディスク装置の位置に物理的に差し替える再構成の作業を促すメッセージ出力などを行う。このホストコンピュータ5におけるメッセージ出力を受けてユーザあるいは保守要員は、ステップS8で、故障ディスク装置となっているデータディスク装置2-3を取り外し、データ復元の済んだ予備ディスク装置4を故障ディスク装置の位置に差し替える位置の移動を行う。

【0117】予備ディスク装置4の故障ディスク装置の位置への移動が済むと、ステップS9で、プロセッサ7は移動後のアレイ構成テーブル12の更新を図18

(E)に示すように行う。勿論、ディスクアレイ装置1から取り外された故障ディスク装置は点検修理が行われることになる。また、空きとなった予備ディスク装置4の位置には別の正常なディスク装置あるいは処理が済んだ故障ディスク装置が実装されることになる。

10. 人的介入を伴わない再構成

図19は何らかに原因によりデータ復元が必要と判断されて、予備ディスク装置へのデータ復元が済んだ後の人的介入によるディスク装置の移動を必要としない再構

成処理を示している。

【0118】図19にあっては、ステップS1で、何らかの原因により例えばデータディスク装置2-3のデータ復元が必要となった場合の処理を示している。この場合、ステップS2でまず復元前のアレイ構成テーブル12をプロセッサ7が参照する。この場合のアレイ構成テーブル12は、例えば図20(B)に示す内容を有する。図20(B)において、アレイ構成テーブル12上で各ディスク装置は「a b y, a b」で表わされる。先頭の「a, b」は物理的な位置を示すランク番号aとポート番号bである。次のyはデータディスクD, 冗長ディスクP, 予備ディスクH, 故障ディスクF, データ復元中ディスクRとなる動作状態を示す種別情報である。最後の「a, b」は論理的な位置を示すランク番号aとポート番号bである。

【0119】このことから、図20(B)のデータ復元前のアレイ構成テーブル12にあっては、ランク番号#0に属する6台のディスク装置について、ポート番号#0~#3のディスク装置については「00D00」~「03D03」が登録され、またポート番号#4のディスク装置については「04P04」が登録され、更にポート番号#5のディスク装置については「05H」が登録されており、前半の物理的な位置と後半の論理的な位置の値は共に等しく、物理的な位置と論理的な位置が1対1に対応している。

【0120】このようなアレイ構成テーブル12をもつ各ディスク装置につき、ステップS3で予備ディスク装置4に対するデータ復元が必要となったデータディスク装置2-3のデータの復元処理を開始する。データ復元中にあっては、ステップS4でアレイ構成テーブルの更新を行う。この場合のアレイ構成テーブル12は、図20(C)に示すように、データ復元が必要となったポート番号#3のディスク装置2-3の登録内容を「03F03」として故障ディスク装置に設定し、またポート番号#5のディスク装置を「05R05」としてデータ復元中のディスク装置であることを示す。

【0121】ステップS5で予備ディスク装置4へのデータ復元が終了すると、ステップS6で予備ディスク装置4の論理的な位置の移動を行い、ステップS7で、この論理的な位置移動に伴うアレイ構成テーブル12の更新を行う。即ち、図20(D)に示すように、ポート番号#5のデータ復元が済んだ予備ディスク装置4について「05D02」に更新し、ポート番号#3の故障ディスク装置のもっていた論理的な位置を示す番号「03」に変更する。

【0122】このため、それ以降のホストコンピュータ5からのリード要求およびライト要求については、ポート番号#5に接続されているデータ復元の済んだ予備ディスク装置4が論理的な位置「02」をもつデータディスク装置として扱われる。また故障ディスク装置となっ

たポート番号#3のデータディスク装置2-3は論理的な位置から除外され、故障ディスク装置を取り外し、新たに正常なディスク装置を実装した際、アレイ構成テーブル12のポート番号#3の位置に予備ディスク装置としての「03H05」の登録が行われ、異なった物理的な位置で予備ディスク装置4として機能するようになる。

【0123】勿論、図19によってはステップS7でデータ復元終了後のアレイ構成テーブルの更新が済むと、ステップS8でデータ復元終了をホストコンピュータ5に通知し、通常のホストコンピュータ5からのコマンド要求に基づく入出力処理に戻る。尚、上記の各実施例にあっては、ディスクアレイ装置1のプロセッサ7において配下のディスク装置でデータ復元が必要なことを認識した場合の予備ディスク装置に対するデータ復元の起動を行っているが、ディスクアレイ装置1からホストコンピュータ5に対しデータ復元が必要なディスク装置が認識されたことを通知し、ホストコンピュータ5からのコマンドにより予備ディスク装置4に対するデータ復元を開始するようにしてもよい。

【0124】この処理は例えば図5のフローチャートのステップS23からS24の処理に移行する部分に示すように上位装置から行われる。この点は、図7、図9、図10、図11、図13、図15、図17及び図19のフローチャートについても同様である。このように予備ディスク装置4に対するデータの復元をホストコンピュータ5側で管理することで、ディスクアレイ装置1側の処理負担を軽減すると同時に、ディスクアレイ装置1に対する入出力要求の空き時間を効率的に利用した予備ディスク装置4へのデータ復元処理を可能とする。

【0125】また本発明は1ランクに6台のディスク装置を設けた場合を例にとっているが、ランク数および1ランク当りのディスク装置の台数は必要に応じて適宜に定めることができる。また1ランクに設ける予備ディスク装置を1台としているが、複数台設けてもよい。また1ランクに1台、予備ディスク装置を設けず、複数ランク当り1台の予備ディスク装置を設けてもよい。

【0126】更にまた、上記の実施例はホストブロックデータをビットまたはバイト単位に分配結合して、ポートに並列接続された複数のディスク装置を並列動作し、またパリティデータを格納するディスク装置を特定ポートに固定したRAID3のディスクアレイ制御形態を例にとっているが、セクタ単位にデータのリード、ライトを行い、セクタごとにパリティデータを格納するディスク位置が変化するRAID5のディスクアレイ制御形態についても、例えばパリティグループを構成する全ディスク装置に対しセクタデータを並列的に読み書きするような場合については、そのまま適用することができる。

【0127】

【発明の効果】以上説明してきたように本発明によれ

ば、次の効果が得られる。まずタイマによりリード動作実行時のデータ到達時間を監視し、パリティグループの中の1台のディスク装置からのデータ転送が遅れた場合には、既に得られているデータおよびパリティデータからデータを復元して上位装置に返送することで、上位装置に対する転送速度を向上できる。

【0128】また訂正不可能なデータブロックについて、同一パリティグループに属する他のディスク装置のデータから訂正不可能なデータを復元して交代領域に割り付けるため、訂正ブロックを検出したディスク装置を故障ディスク装置として扱わなくてもよく、ディスク装置の信頼性を向上することができる。更に、交代領域が全て使い尽くされたり残り数がなくなったディスク装置のデータを予備ディスク装置に復元して故障ディスク装置として扱うことで、交代領域のアクセスに伴う応答時間の遅れを防止し、ディスクアレイ装置の処理性能を維持することができる。

【0129】更にまた、同期回転を維持できなくなったディスク装置についても、予備ディスク装置にデータを復元して故障ディスク装置として扱うことで、同期回転が維持できなくなった装置が論理ディスク装置を構成するディスク装置の中に存在することで処理性能が低下してしまうことを防止できる。

【0130】更にまた、同一トラックまたは同一シリンダの交代領域ブロック数が所定値を越えた場合に、予備ディスク装置にデータを復元して故障ディスク装置として扱うことで、交代領域が複数トラックや複数シリンダ位置に及ぶことによるアクセス性能の低下を防止できる。更にまた、ディスク装置のエラー回数が閾値に達したときに予備ディスク装置へのデータ復元を行って故障ディスク装置として扱うことで、故障の危険性の高いディスク装置を論理ディスク装置を構成するディスク装置の中から除外し、信頼性を向上できる。

【0131】一方、何らかの原因によりデータ復元が必要となって、予備ディスク装置に対するデータ復元が終了した後の予備ディスク装置の故障ディスク装置との置き替えについて、置き替えを促すメッセージ出力を自動的に行うことで、物理的な移動を伴う人的な再構成を確実にできる。更に、予備ディスク装置の論理的な位置をアレイ構成テーブル上で書き替えるだけで、物理的な故障ディスク装置と予備ディスク装置の移動を必要とすることなく簡単にディスクアレイの再構成が実現できる。

【0132】

【図面の簡単な説明】

【図1】本発明の原理説明図

【図2】本発明のハードウェア構成と機能を示した実施例構成図

【図3】タイマ監視によるデータ復元処理を示したフロー

ーチャート

【図4】データ復元を伴う交代割付処理を示したフローチャート

【図5】交代領域オーバーフローに伴う予備へのデータ復元処理を示したフローチャート

【図6】図5の続きを示したフローチャート

【図7】交代領域オーバーフローの予測に基づくデータ復元処理を示したフローチャート

【図8】図8の続きを示したフローチャート

【図9】同期回転異常時の予備へのデータ復元処理を示したフローチャート

【図10】同期回転異常時の予備へのデータ復元処理の他の実施例を示したフローチャート

【図11】障害発生予測と予備へのデータ復元処理を示したフローチャート

【図12】図11の続きを示したフローチャート

【図13】同一トラック又はシリンダ内の交代ブロック数が増加した場合の予備へのデータ復元処理を示したフローチャート

【図14】図13の続きを示したフローチャート

【図15】同一トラック又はシリンダ内の交代ブロック数が増加した場合の予備へのデータ復元処理の他の実施例を示したフローチャート

【図16】図15の続きを示したフローチャート

【図17】人的介入により物理的なディスク装置の移動を伴う再構成処理を示したフローチャート

【図18】図17の処理で更新されるアレイ構成テーブルの内容を示した説明図

【図19】人的介入により物理的なディスク装置の移動を必要としない再構成処理を示したフローチャート

【図20】図19の処理で更新されるアレイ構成テーブルの内容を示した説明図

【図21】従来装置の説明図

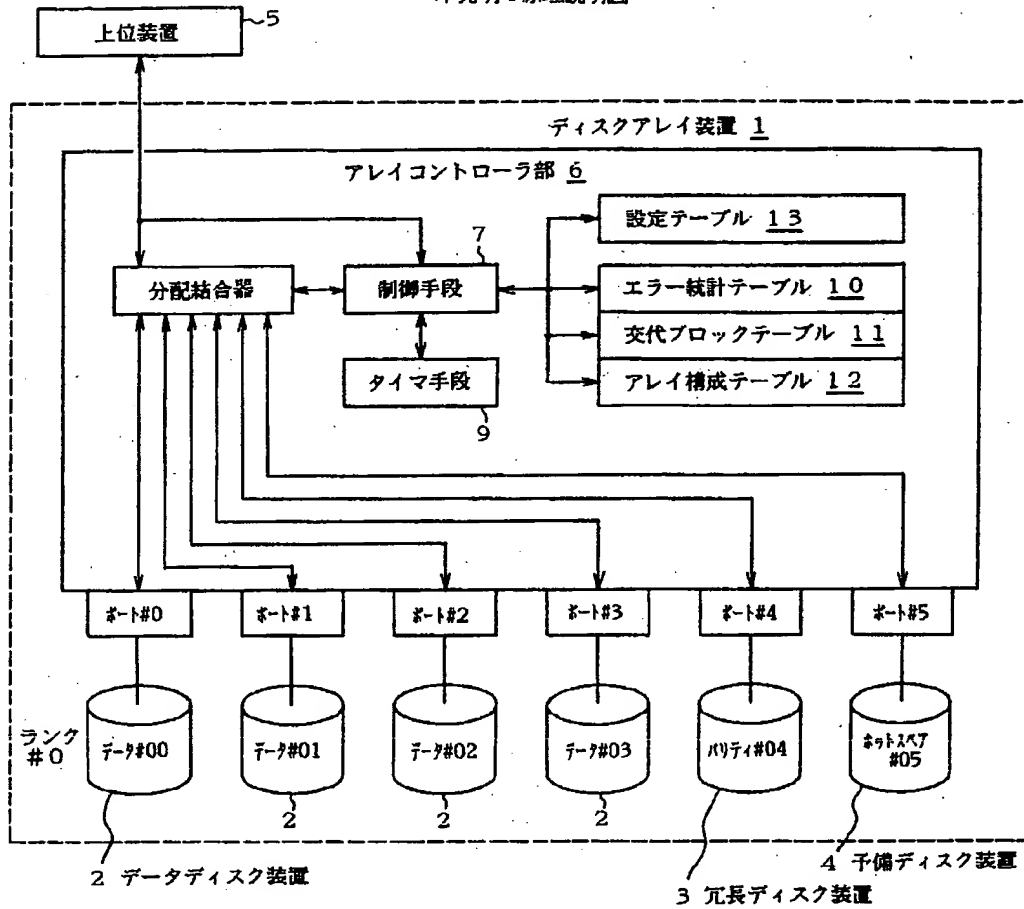
【0133】

【符号の説明】

- 1：ディスクアレイ装置
- 2, 2-0～2-3：データディスク装置
- 3：冗長ディスク装置
- 4：予備ディスク装置
- 5：上位装置（ホストコンピュータ）
- 6：アレイコントローラ部
- 7：プロセッサ（制御手段）
- 8：分配結合器
- 9：タイマ（タイマ手段）
- 10：エラー統計テーブル
- 11：交代ブロックテーブル
- 12：アレイ構成テーブル
- 13：閾値設定テーブル

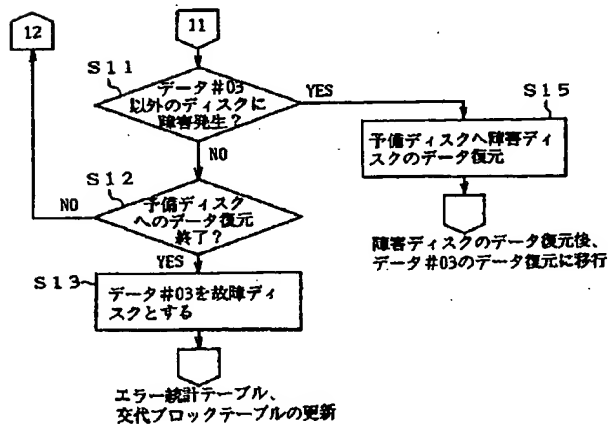
【図1】

本発明の原理説明図



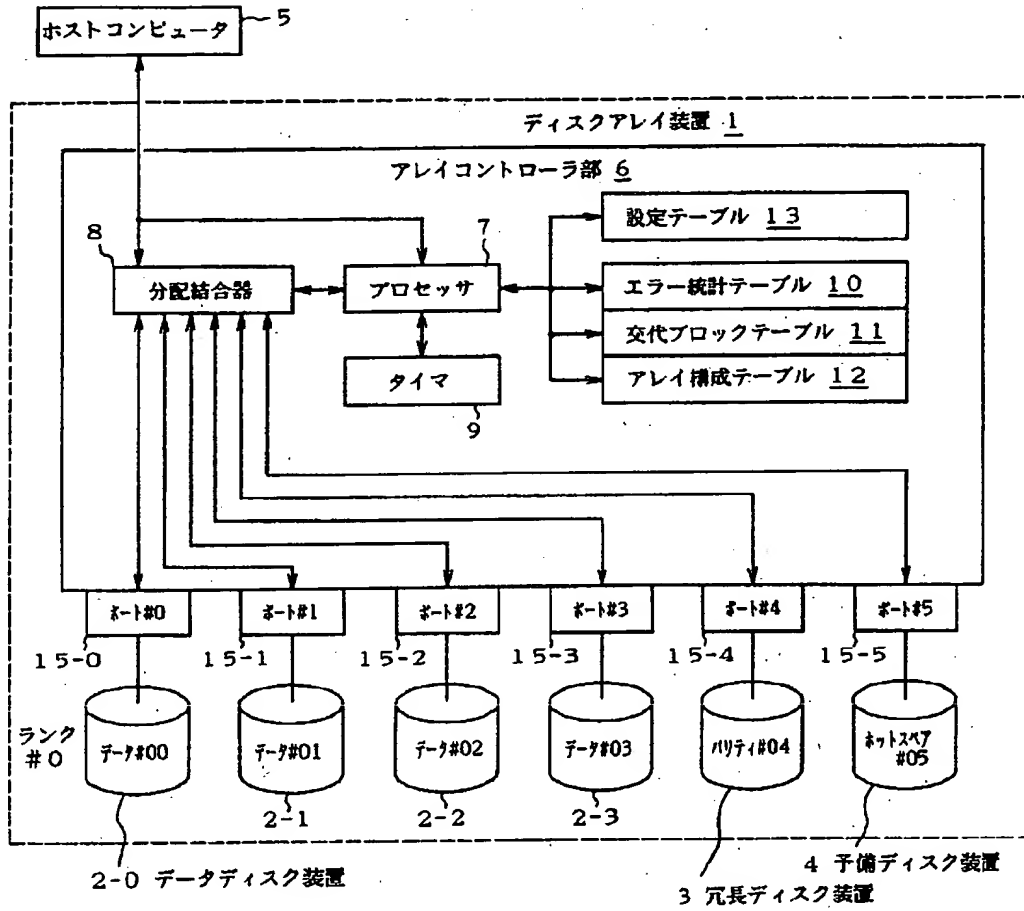
【図16】

図15の続きを示したフローチャート



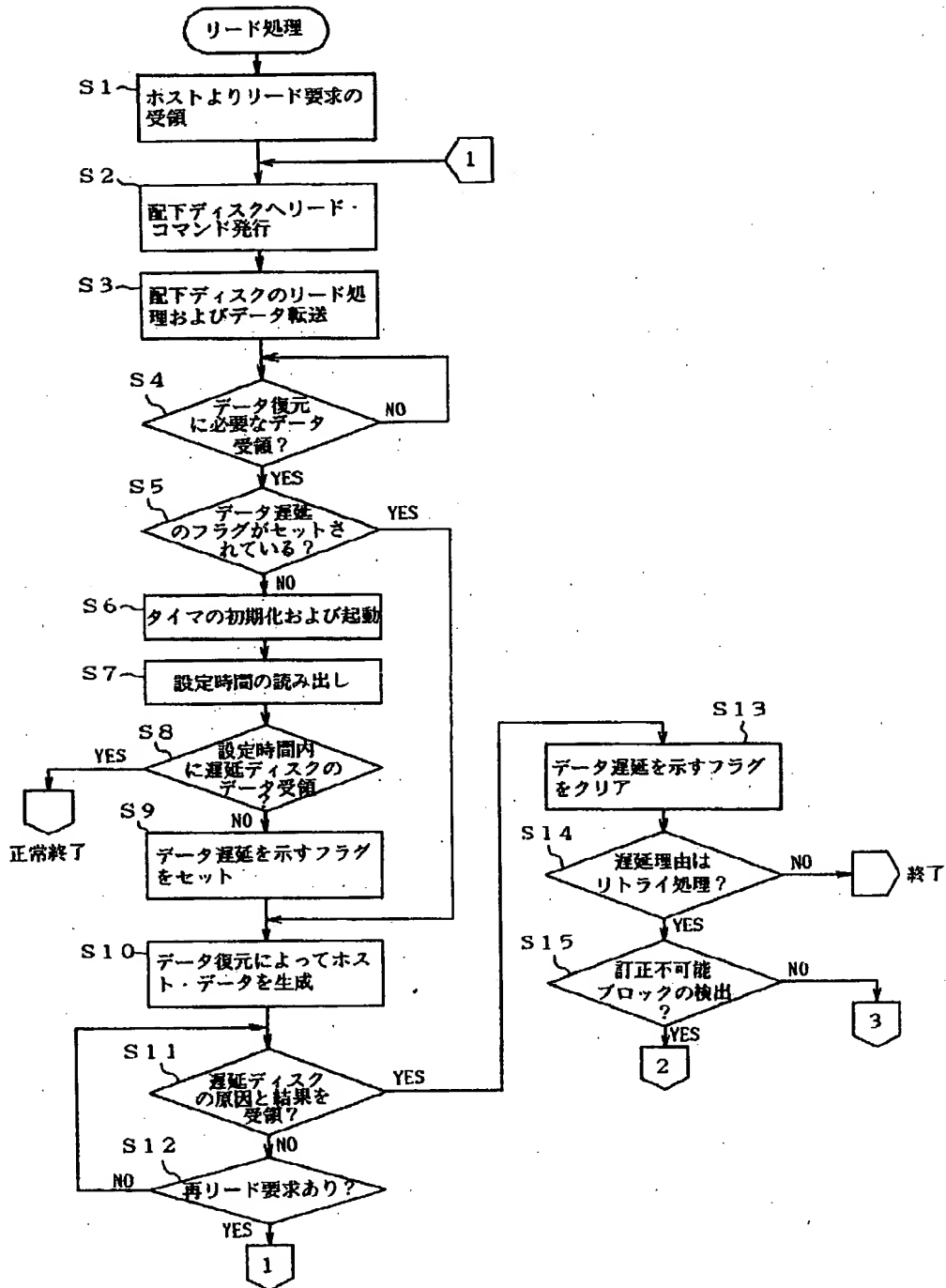
【図2】

本発明のハードウェア構成と機能を示した実施例構成図



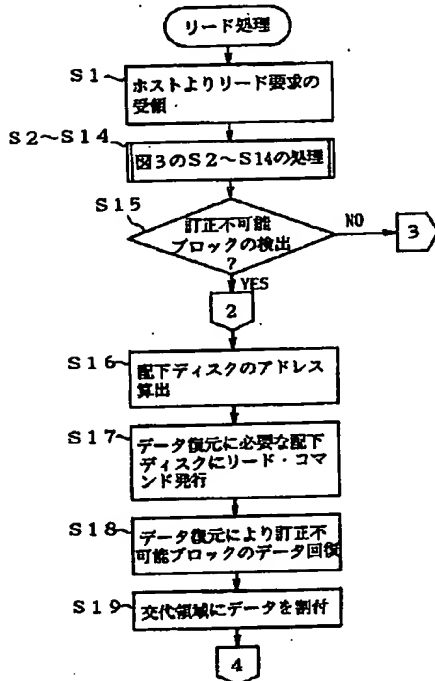
【図3】

タイマ監視によるデータ復元処理を示したフローチャート



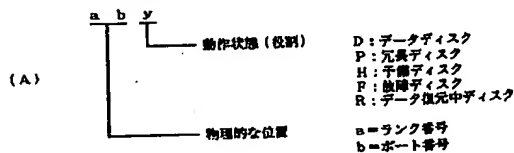
【図4】

データ復元を伴う交代割付処理を示したフローチャート



【図18】

図17の処理で更新されるアレイ構成テーブルの内容を示した説明図



(データ復元前)

	ポート#0	ポート#1	ポート#2	ポート#3	ポート#4	ポート#5
ランク#0	データ#00	データ#01	データ#02	データ#03	パリティ#0	データ#04
	00D	01D	02D	03D	04P	05H

(データ復元中)

	ポート#0	ポート#1	ポート#2	ポート#3	ポート#4	ポート#5
ランク#0	データ#00	データ#01	データ#02	データ#03	パリティ#0	データ#04
	00D	01D	02D	03R	04P	05R

(データ復元後)

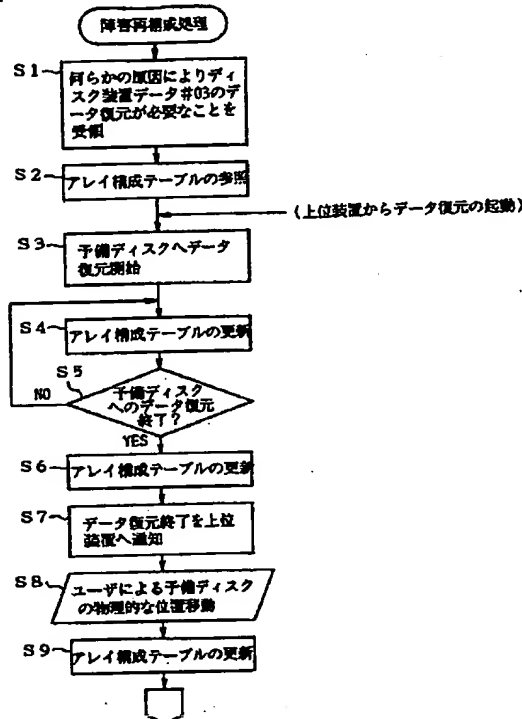
	ポート#0	ポート#1	ポート#2	ポート#3	ポート#4	ポート#5
ランク#0	データ#00	データ#01	データ#02	データ#03	パリティ#0	データ#04
	00D	01D	02D	03D	04P	05D

(ディスク移動後)

	ポート#0	ポート#1	ポート#2	ポート#3	ポート#4	ポート#5
ランク#0	データ#00	データ#01	データ#02	データ#03	パリティ#0	
	00D	01D	02D	03D	04P	

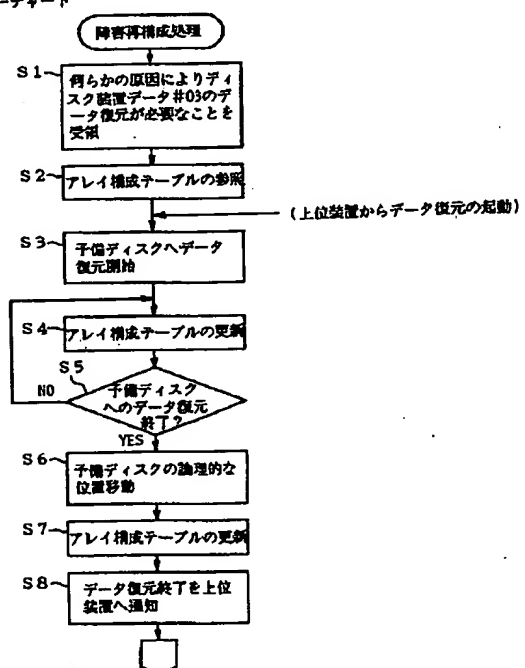
【図17】

人的介入により物理的なディスク装置の移動を伴う再構成処理を示したフローチャート



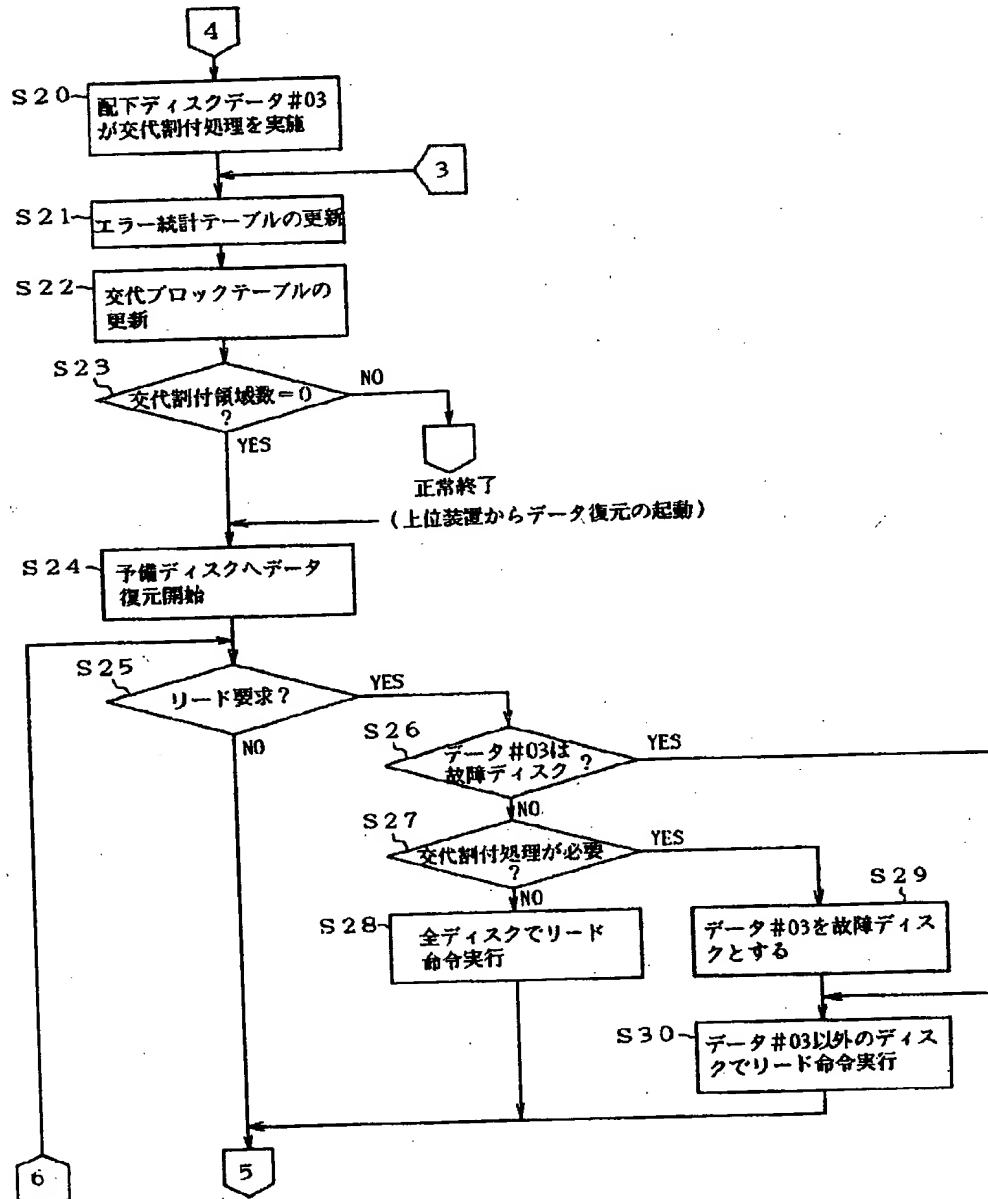
【図19】

人的介入により物理的なディスク装置の移動を必要としない再構成処理を示したフローチャート



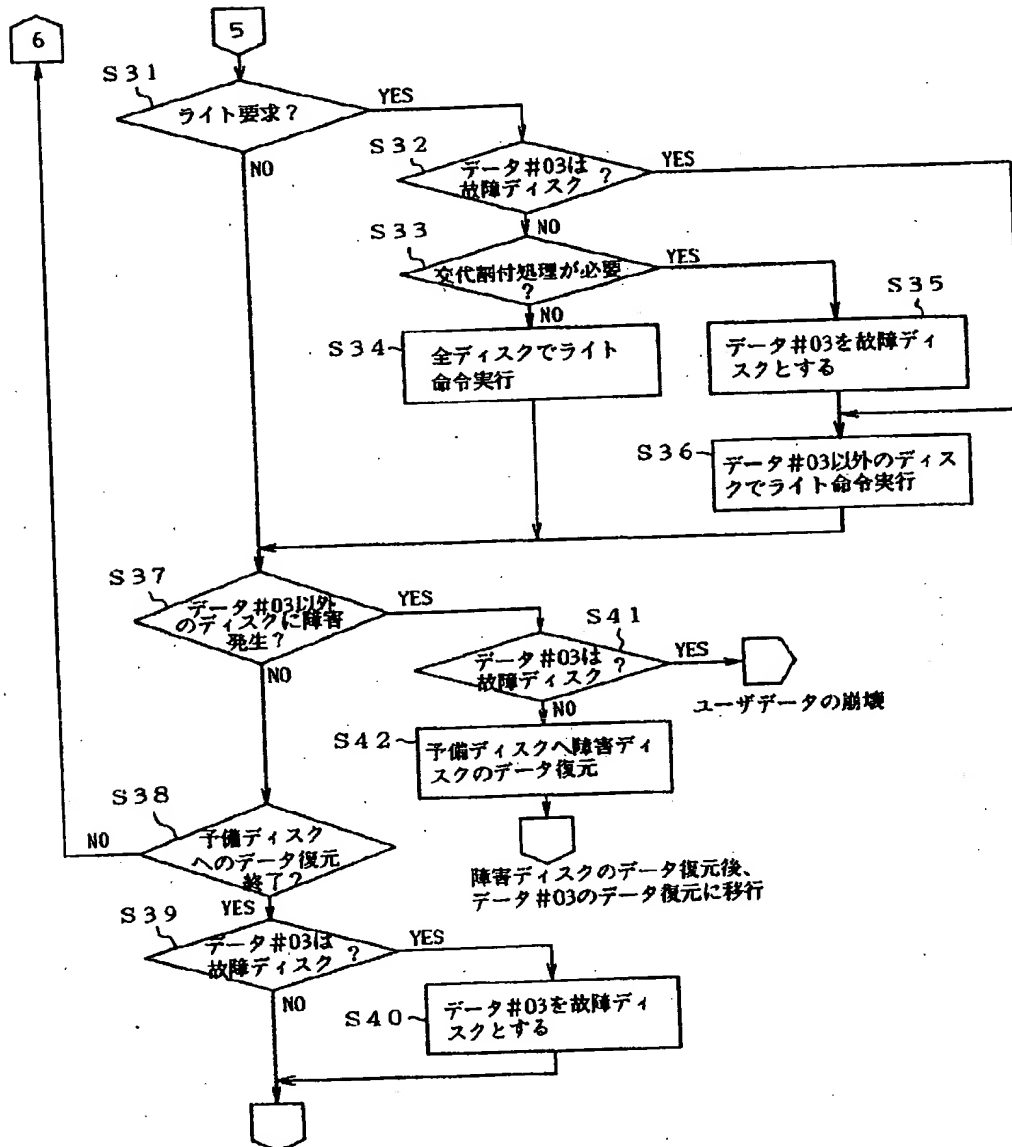
【図5】

交代領域オーバーフローに伴う予備へのデータ復元処理を示したフローチャート



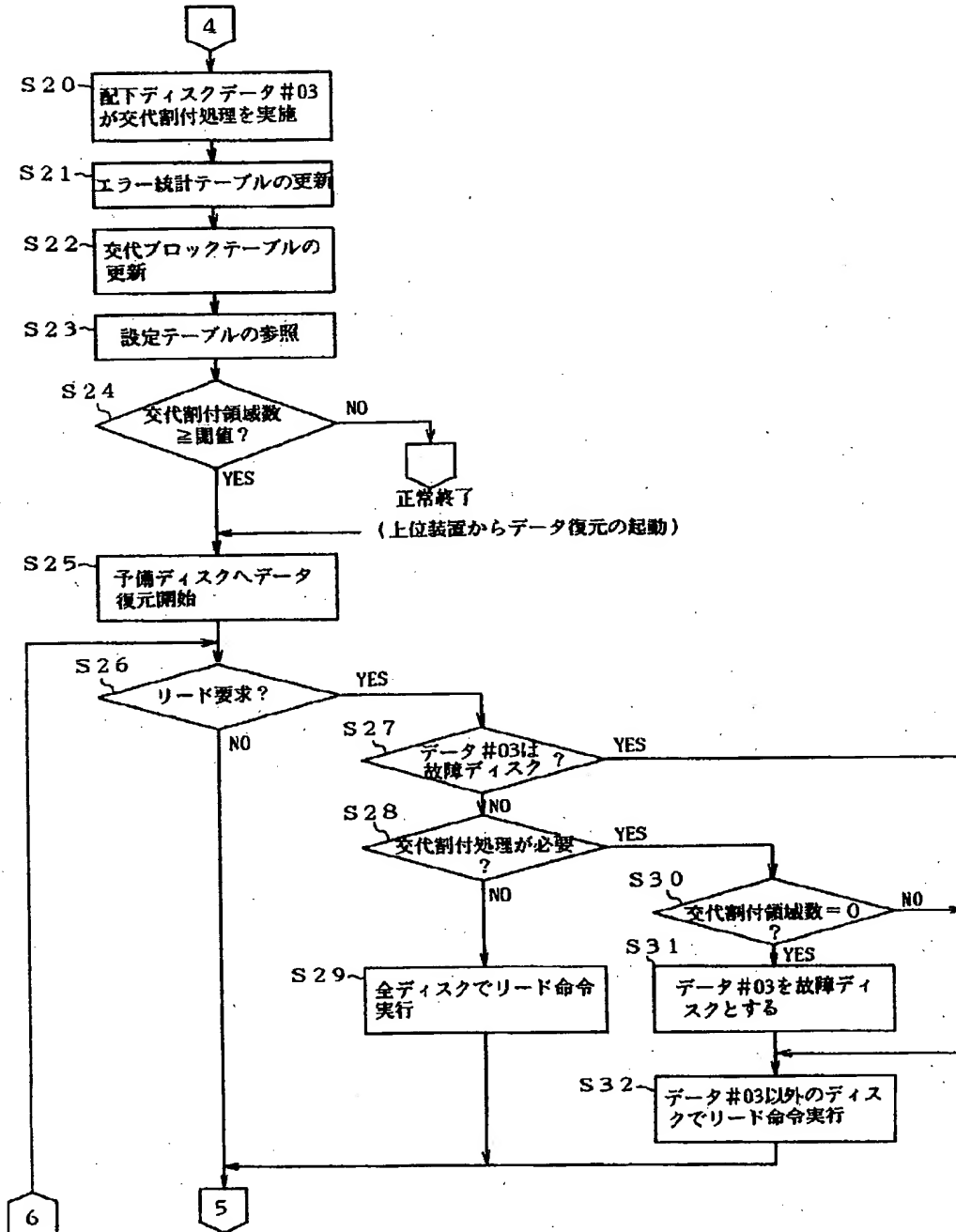
【図6】

図5の続きを示したフローチャート



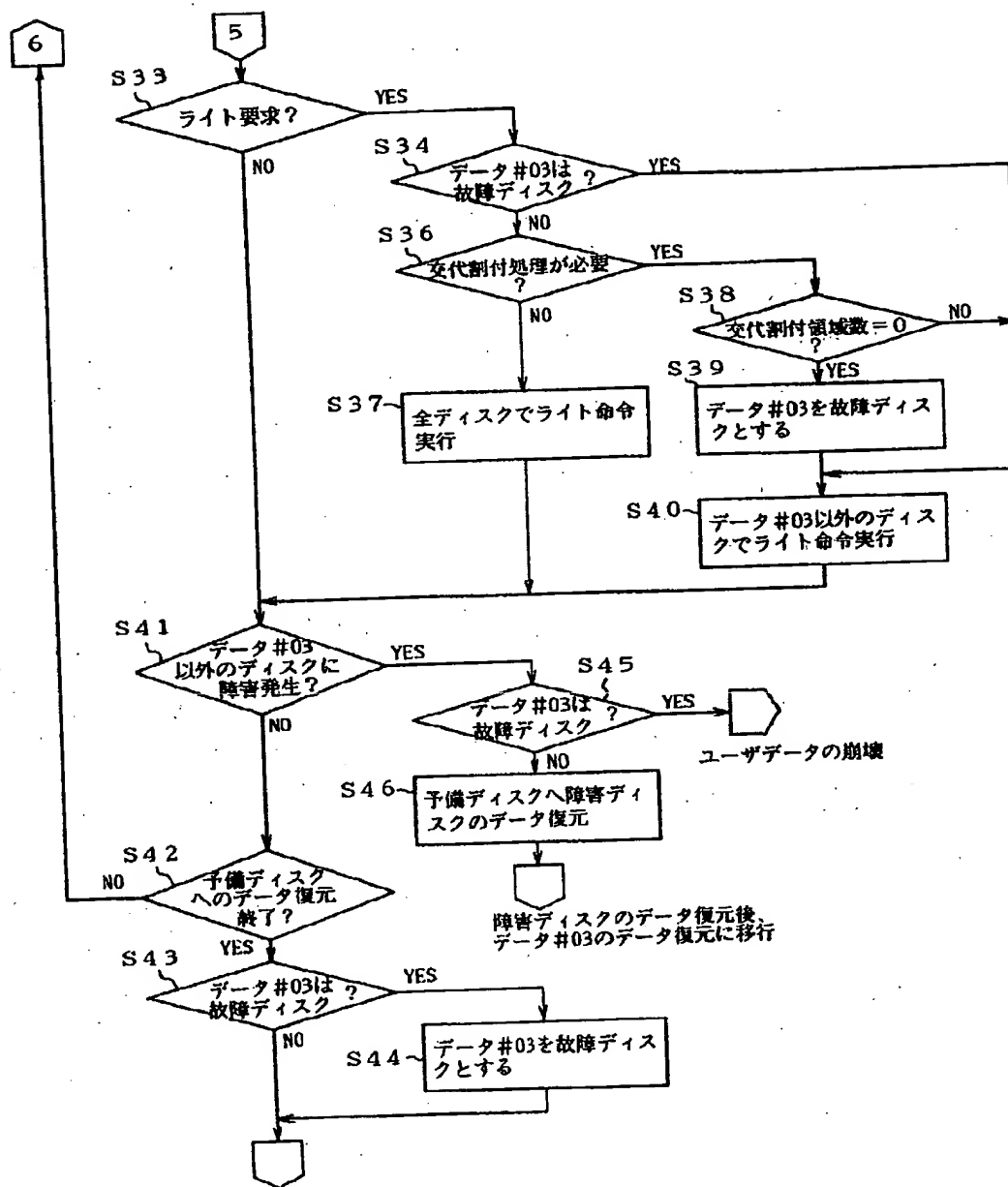
【図7】

交代領域オーバーフローの予測に基づくデータ復元処理を示したフローチャート



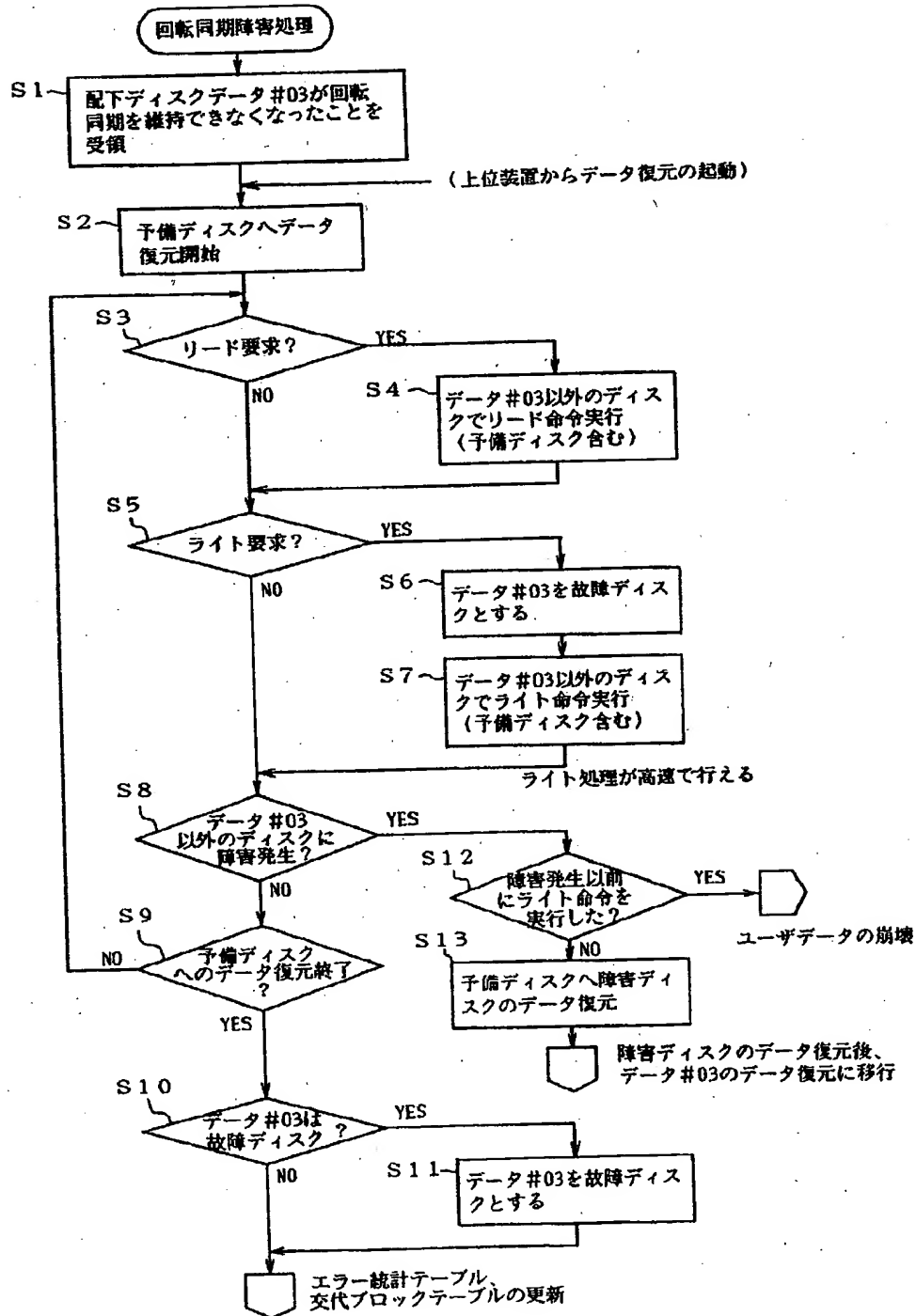
【図8】

図7の続きを示したフローチャート



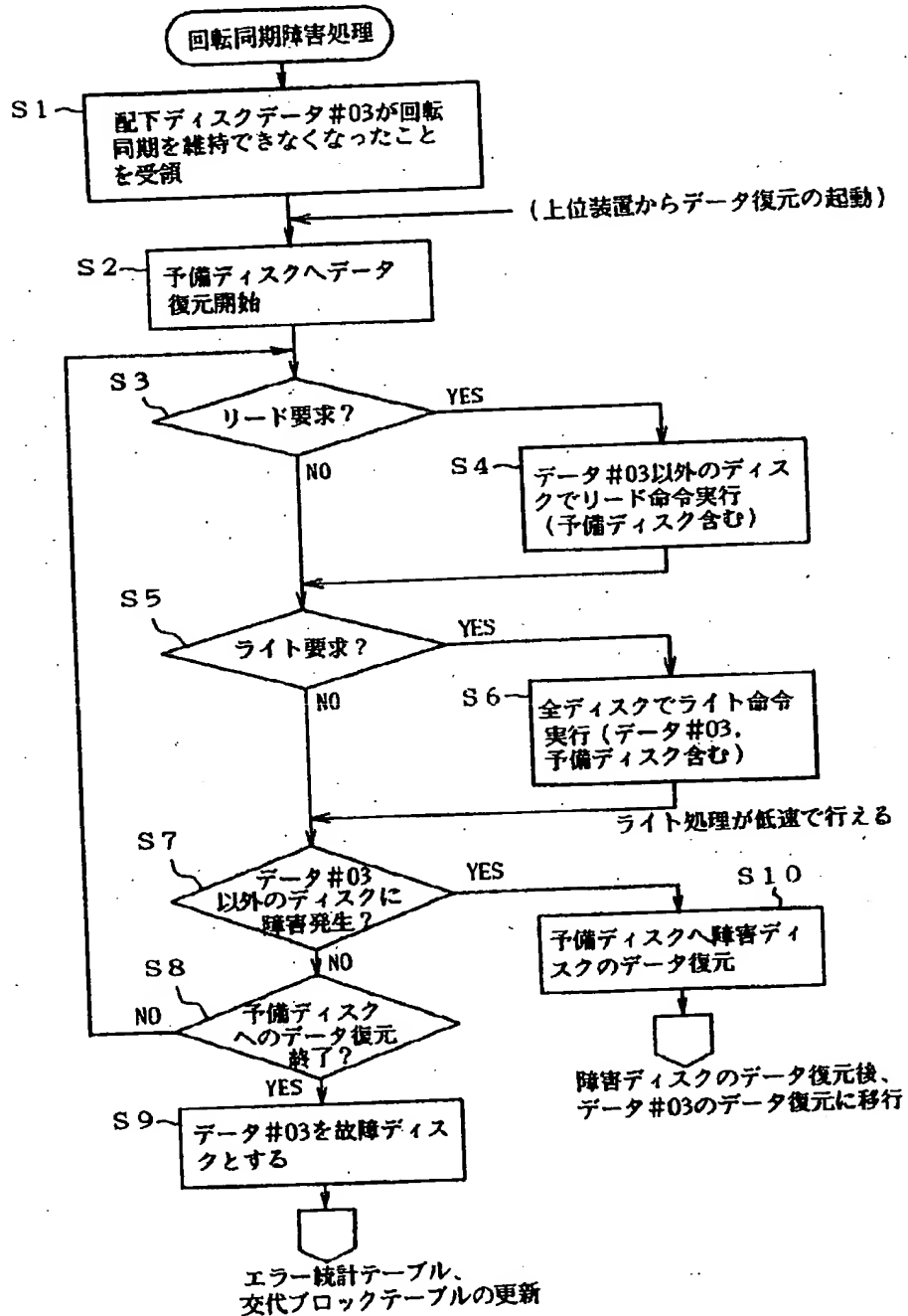
【図9】

同期回転異常時の予備へのデータ復元処理を示したフローチャート



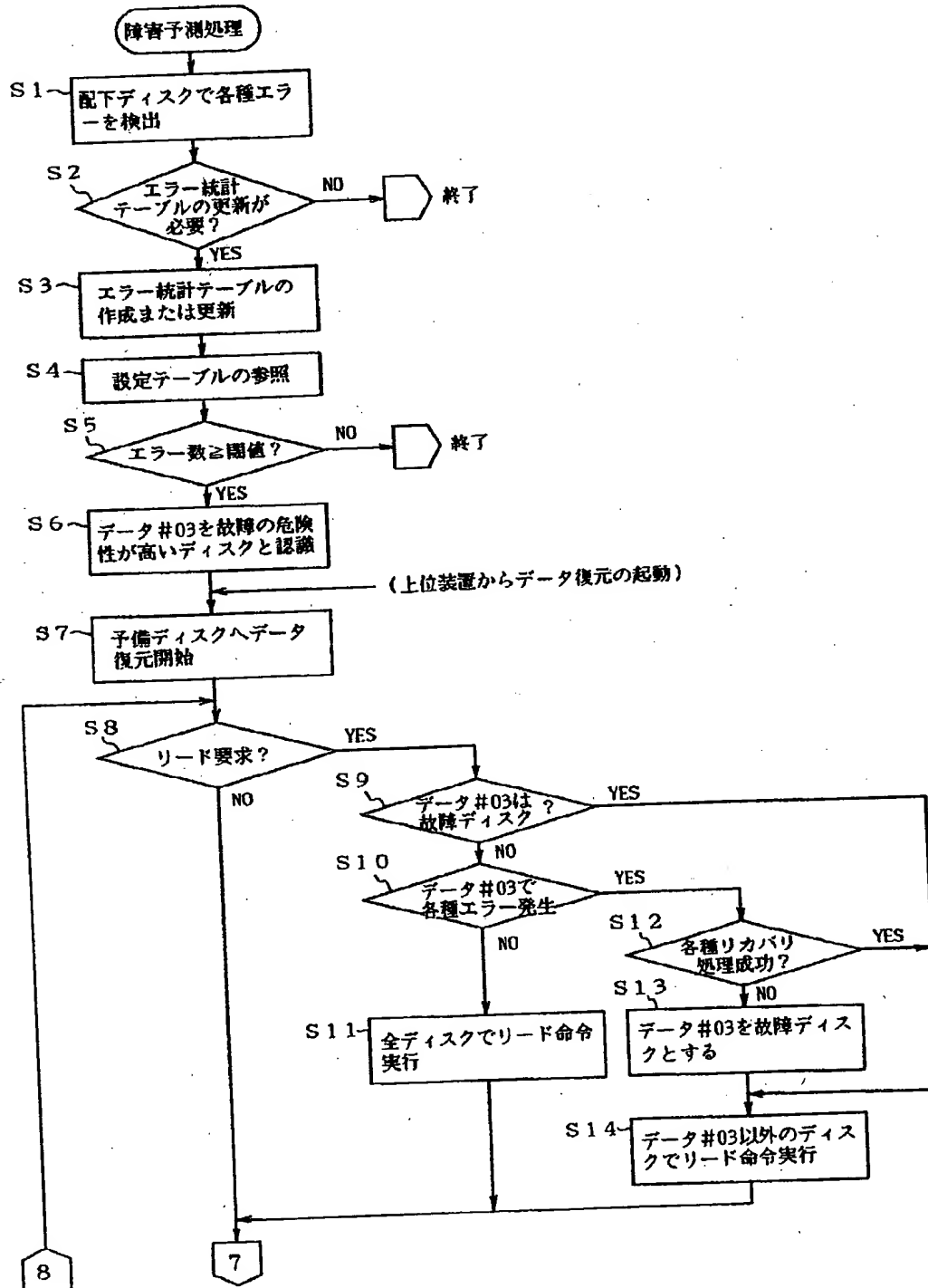
【図10】

同期回転異常時の予備へのデータ復元処理の他の実施例を示したフローチャート



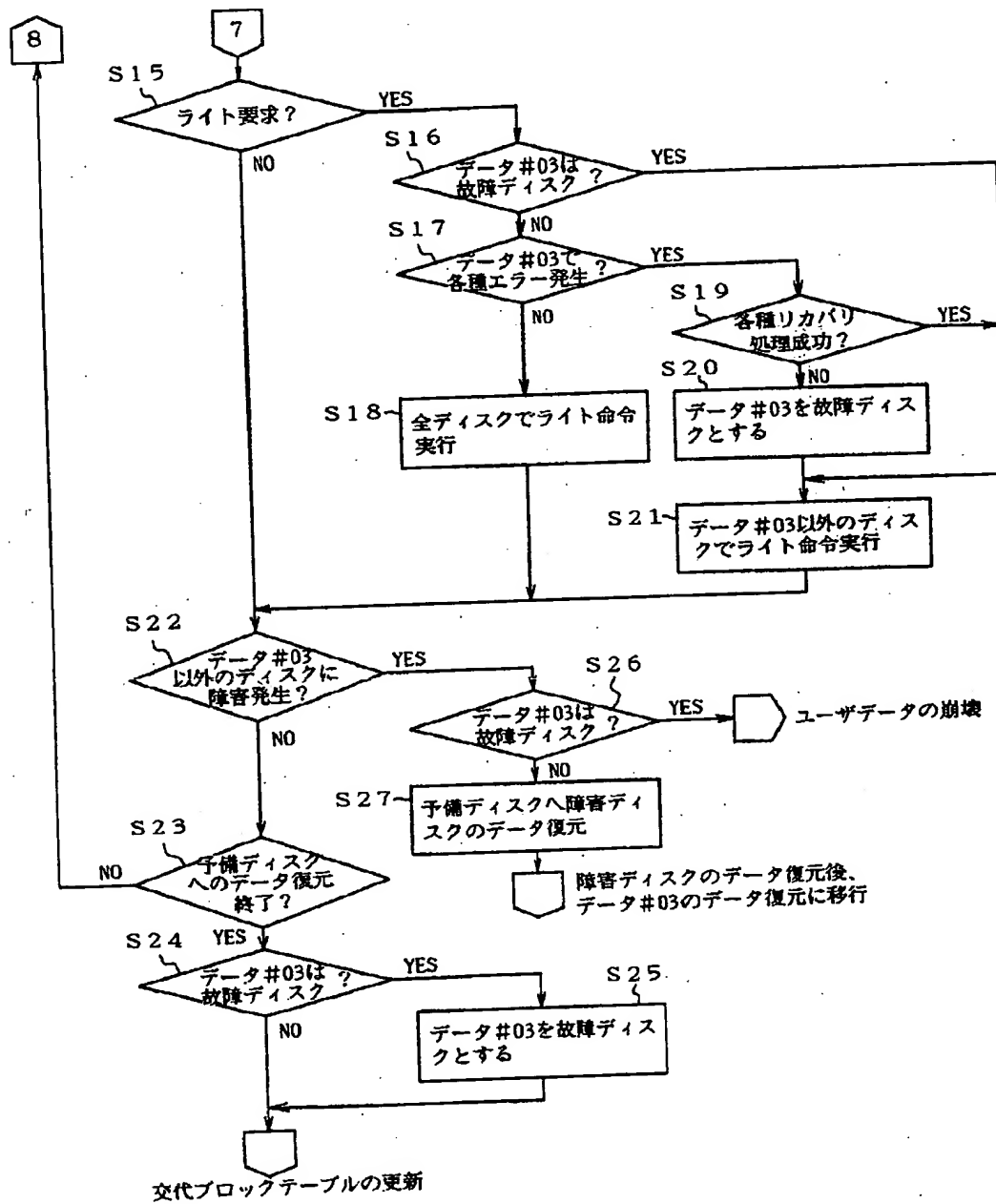
【図11】

障害発生予測と予備へのデータ復元処理を示したフローチャート



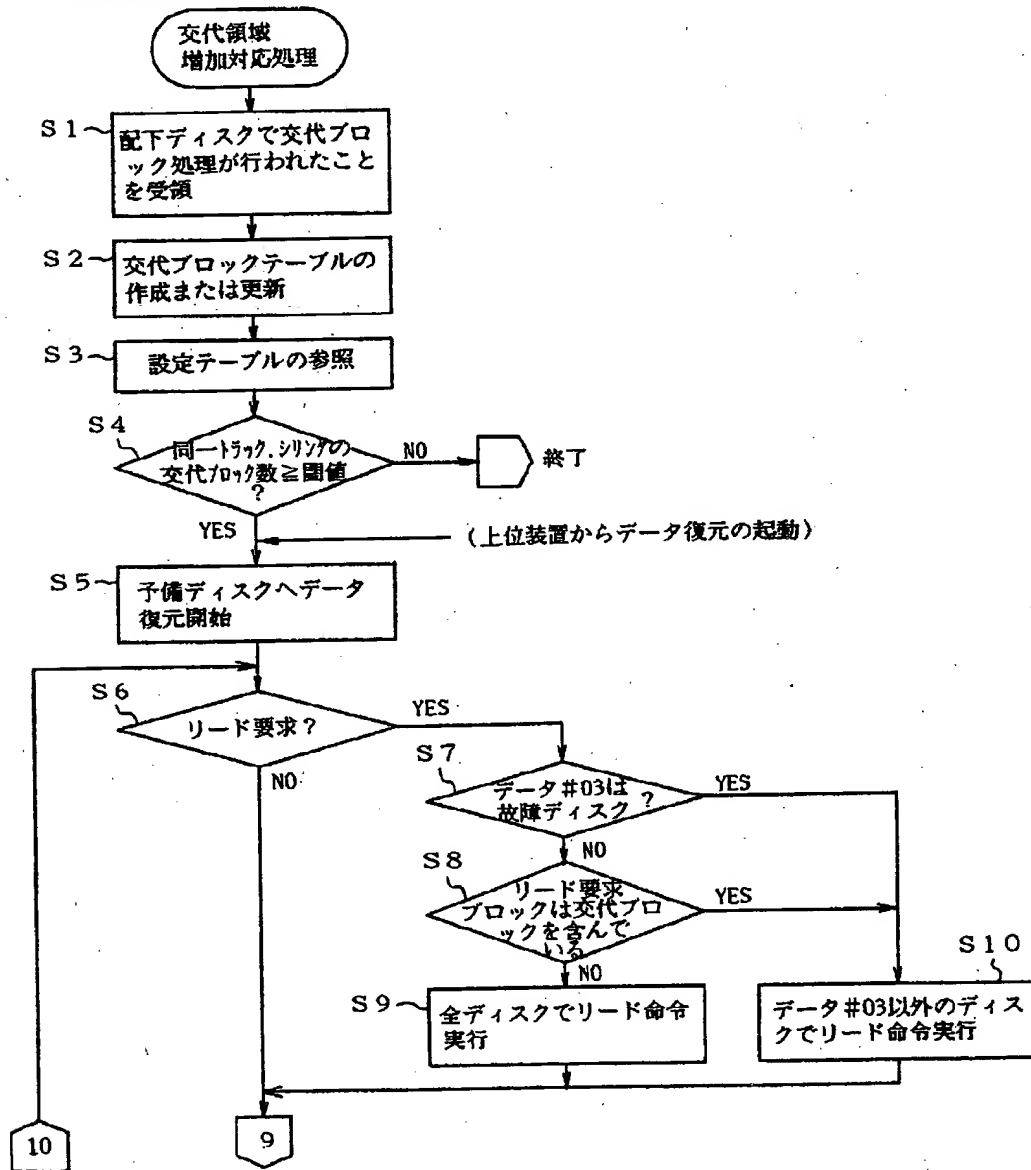
【図12】

図11の続きを示したフローチャート



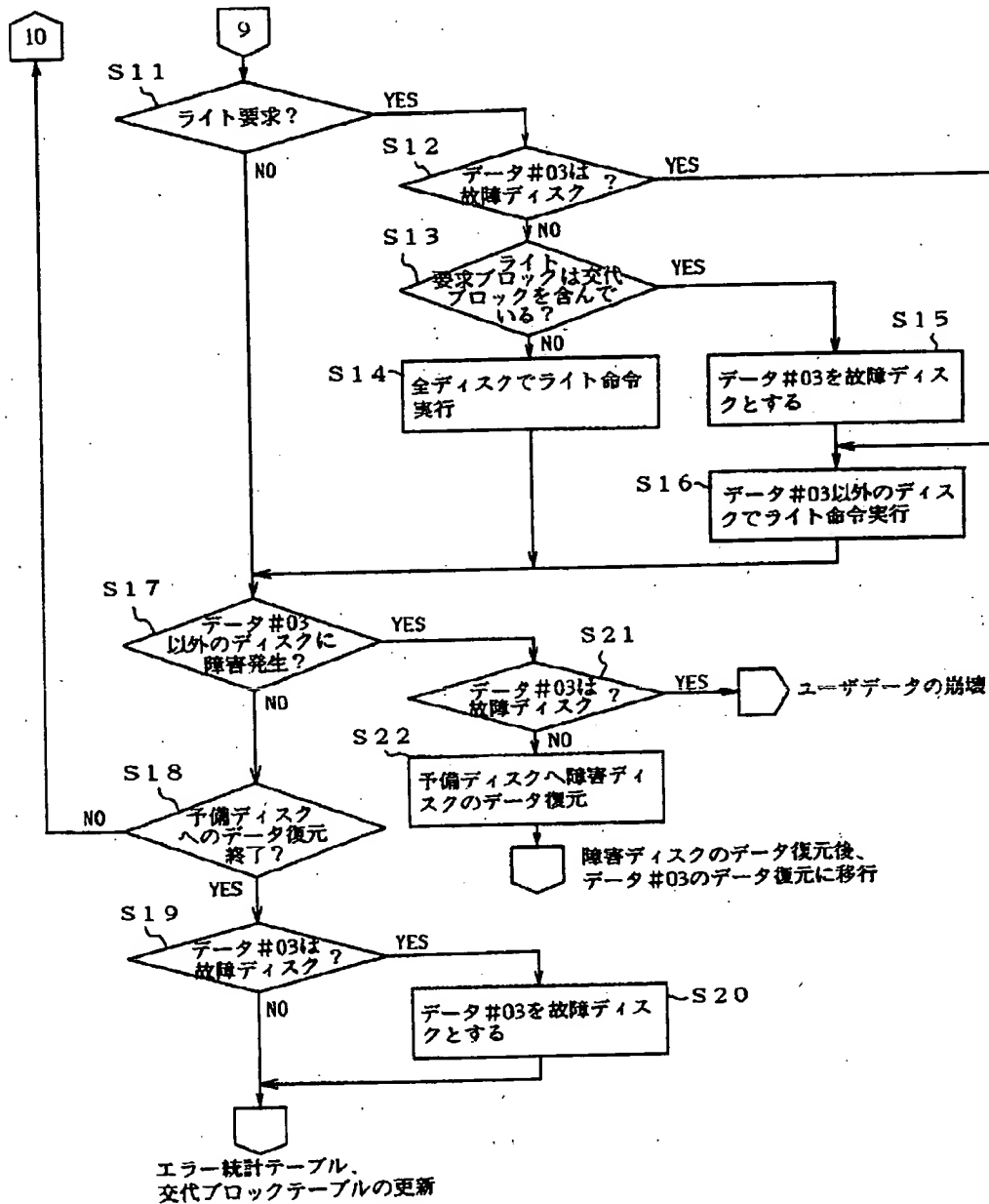
【図13】

同一トラック又はシリンダ内の交代ブロック数が増加した場合の予備へのデータ
復元処理を示したフローチャート



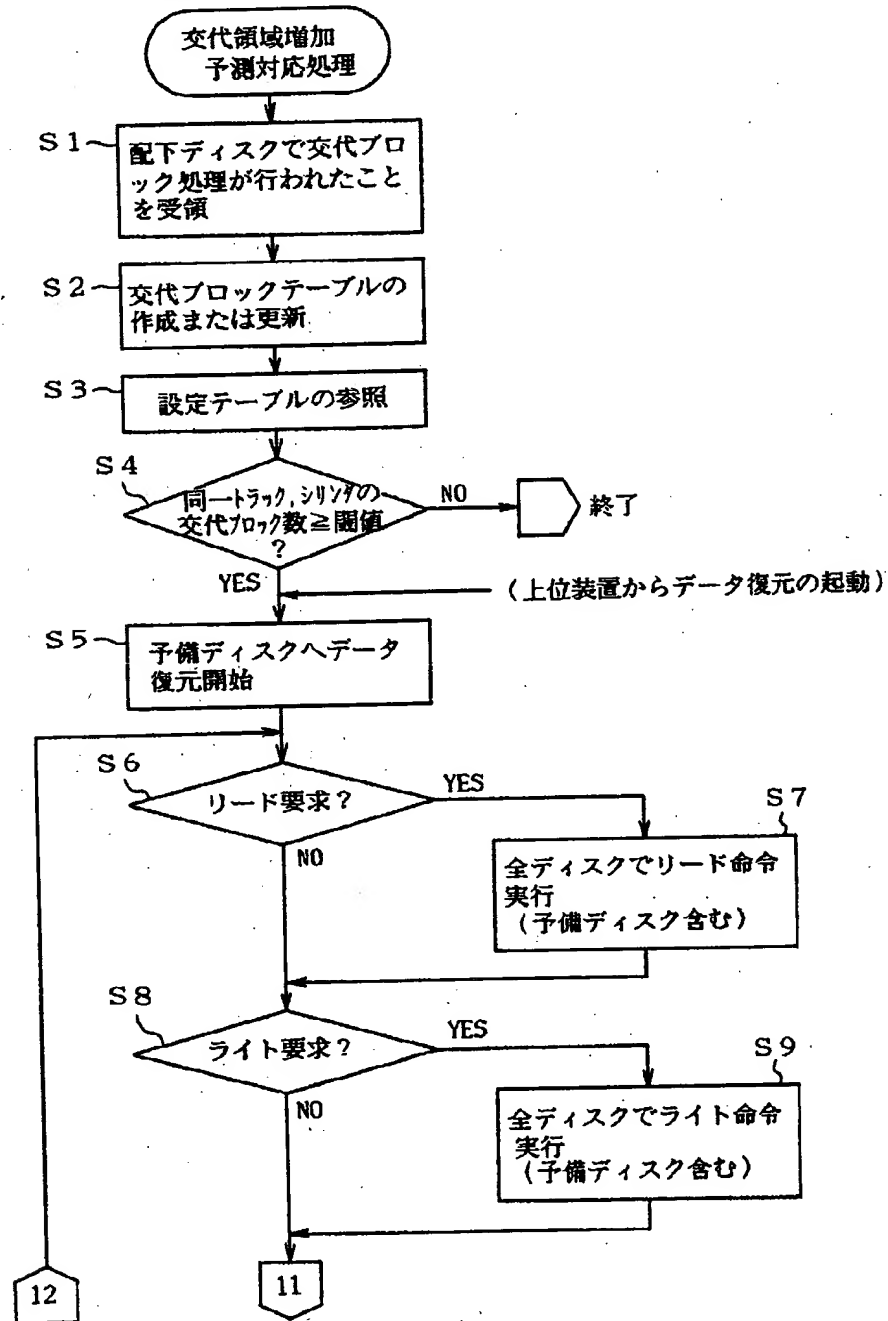
【図14】

図13の続きを示したフローチャート



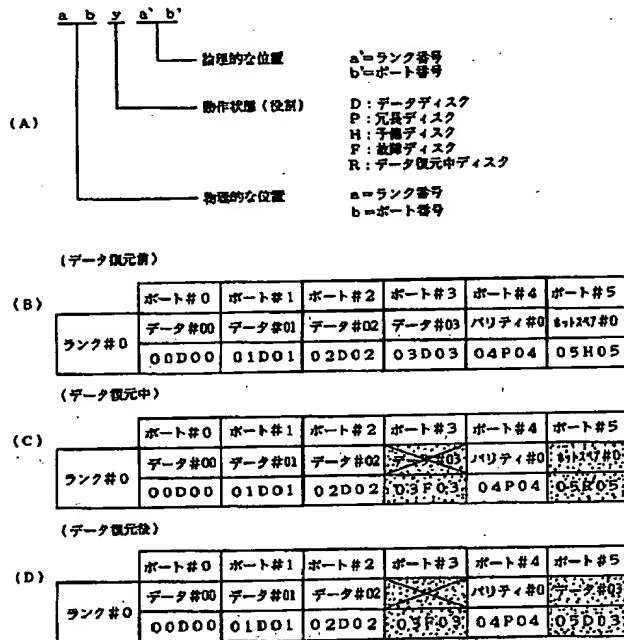
【図15】

同一トラック又はシリンダ内の交代ブロック数が増加した場合の予備へのデータ復元処理の他の実施例を示したフローチャート



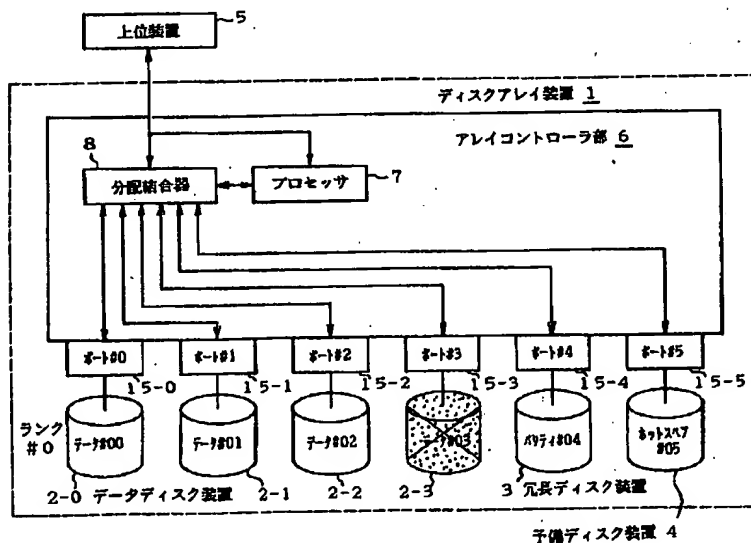
【図20】

図19の処理で更新されるアレイ構成テーブルの内容を示した説明図



【図21】

従来装置の説明図



THIS PAGE BLANK (USPTO)